

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Commissioner
 US Department of Commerce
 United States Patent and Trademark
 Office, PCT
 2011 South Clark Place Room
 CP2/5C24
 Arlington, VA 22202
 ETATS-UNIS D'AMERIQUE
 in its capacity as elected Office

Date of mailing (day/month/year) 11 December 2000 (11.12.00)	
International application No. PCT/US00/09723	Applicant's or agent's file reference BC1003PCT
International filing date (day/month/year) 12 April 2000 (12.04.00)	Priority date (day/month/year) 12 April 1999 (12.04.99)
Applicant HARDER, Patricia, A. et al	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:

30 October 2000 (30.10.00)

☐ in a notice effecting later election filed with the International Bureau on:2. The election ☒ was☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer F. Baechler Telephone No.: (41-22) 338.83.38
---	---

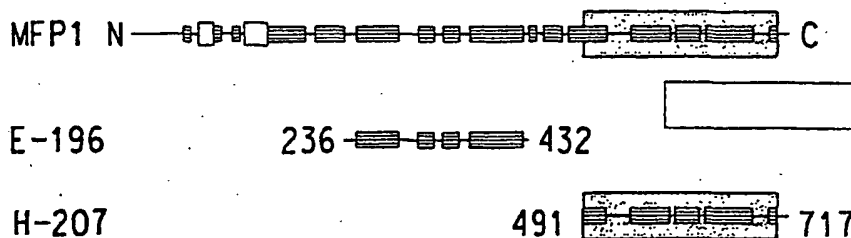
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C07K 14/00	A2	(11) International Publication Number: WO 00/61615 (43) International Publication Date: 19 October 2000 (19.10.00)
(21) International Application Number: PCT/US00/09723 (22) International Filing Date: 12 April 2000 (12.04.00) (30) Priority Data: 60/128,900 12 April 1999 (12.04.99) US (71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): HARDER, Patricia, A. [US/US]; 712 West 34th Street, Wilmington, DE 19802 (US). MEIER, Iris [DE/US]; 4467 Masters Drive, Columbus, OH 43220 (US). (74) Agent: FELTHAM, S., Neil; E.I. Du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).	(81) Designated States: BR, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published Without international search report and to be republished upon receipt of that report.	

(54) Title: HOMOLOGS OF MAR-BINDING FILAMENT-LIKE PROTEIN 1 (MFP1)



(57) Abstract

This invention pertains to nucleic acid fragments encoding proteins that are homologs to the MAR binding filament-like protein 1 (MFP1) from tomato. More specifically, this invention pertains to two tobacco MFP1 genes and MFP1 homologs from corn, soybean, and rice.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LJ	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

2



Applicant's or agent's file reference BC1003		FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/US00/09723	International filing date (day/month/year) 12/04/2000	Priority date (day/month/year) 12/04/1999	
International Patent Classification (IPC) or national classification and IPC C07K14/00			
Applicant E.I. DU PONT DE NEMOURS AND COMPANY et al.			

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of 10 sheets, including this cover sheet.
 - ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 8 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☒ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☒ Certain observations on the international application

Date of submission of the demand 30/10/2000	Date of completion of this report 10.08.2001
Name and mailing address of the International preliminary examining authority:  European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Petri, B  Telephone No. +49 89 2399 7356

Form PCT/IPEA/409 (cover sheet) (January 1994)

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**International application No. **PCT/US00/09723****I. Basis of the report**

1. With regard to the elements of the international application (*Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17):*)
Description, pages:

1-34 as originally filed

Claims, No.:

1-56 as received on 30/10/2000 with letter of 26/10/2000

Drawings, sheets:

1/4-4/4 as originally filed

Sequence listing part of the description, pages:

1-19, as originally filed

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☒ contained in the international application in written form.
- ☒ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/US00/09723

- ☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:
see separate sheet

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

- ☐ the entire international application.
☒ claims Nos. 1-14(partially), 15-56.

because:

- ☒ the said international application, or the said claims Nos. 1-14(partially), 15-56 relate to the following subject matter which does not require an international preliminary examination (*specify*):
see separate sheet
- ☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):
- ☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.
- ☒ no international search report has been established for the said claims Nos. 43-56.
2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:
- ☐ the written form has not been furnished or does not comply with the standard.
☐ the computer readable form has not been furnished or does not comply with the standard.

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/US00/09723

- ☐ restricted the claims.
- ☐ paid additional fees.
- ☐ paid additional fees under protest.
- ☒ neither restricted nor paid additional fees.
2. ☐ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.
3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is
- ☐ complied with.
- ☐ not complied with for the following reasons:
4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:
- ☐ all parts.
- ☒ the parts relating to claims Nos. 1-14(partially).
SEE SEPERATE SHEET
- V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Yes: Claims 1-14
	No: Claims
Inventive step (IS)	Yes: Claims
	No: Claims 1-14
Industrial applicability (IA)	Yes: Claims
	No: Claims none

2. Citations and explanations
see separate sheet**VIII. Certain observations on the international application**

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/US00/09723

1. Reference is made to the following document/s/:

D1: MEIER IRIS ET AL: 'MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA.' PLANT CELL, vol. 8, no. 11, 1996, pages 2105- 2115, XP002139228 ISSN: 1040-4651 cited in the application

2. The present application relates to the cloning of MFP1-like proteins from 4 different plant species (soybean, corn, rice, tobacco). The proteins have been cloned by screening cDNA libraries with probes derived from the respective sequence of the tomato homolog known from D1 (i.e. p7-2 depicted in Seq. Id. No. 9 which corresponds to 1.6 kb of the 3' part of the tomato homolog or p1-3 depicted in Seq. Id. No. 10 which corresponds to 1 kb of the 5' end of the tomato homolog). Sequence comparisons are provided (Table 1, 3). No function beyond that known for LeMFP1 (i.e. the protein known from D1) is disclosed.

3. The claims are directed to several different embodiments:

- 3.1. I. Two full length sequences found in tobacco (Seq. Id. No. 1 and 3) encoding two proteins termed NtMFP1 and NtMFP2 (Seq. Id. No. 2 and 4) (claim 1a, partially).
- II. Portions/fragments of undefined length ("substantial") thereof (claim 1a, partially).
- III. Sequences that are somehow ("substantially") similar thereto (claim 1b, partially).
- IV. Portions/fragments of undefined length ("substantial") of said somehow similar sequences (claim 1b).
- V. Nucleic acid molecules of completely undetermined length that hybridize to said sequences.
- VI. Nucleic acid molecules of completely undetermined length that hybridize to either full length sequence (Seq. Id. No. 1 or 3) or particular portions thereof

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/US00/09723

(i.e. Seq. Id. No. 14 and 15).

VII. Nucleic acid molecules which are complementary to any of the previously defined sequences.

Dependent on the such defined products, claims directed to polypeptides, chimeric genes, host cells, and methods are claimed (claim 2-14).

3.2 An analogous set of claims is directed to a further MFP1 protein found in soybean (Seq. Id. No. 20)(claims 15-28),

3.3 and further MFP1 proteins found in corn (Seq. Id. No. 22)(claims 29-42)

3.4 and rice (Seq. Id. No. 24)(claims 43-56).

Re Item I**Basis of the opinion**

4. The amendments filed with the International Bureau under Article 34 PCT appear to be in accordance with the requirements of Article 34(2)(b) PCT.

Re Item III**Non-establishment of opinion with regard to novelty, inventive step and industrial applicability**

5. The international search was restricted to the invention groups 1-3 as covered by the claims as specified on form ISA210-1, Box II. Consequently no examination will be carried out for those parts of the application which were not searched (Rule 66.1(e) PCT), namely claims 43-56 relating to Seq. Id. No. 23 and 24.

6. In reply to the invitation to restrict the claims or to pay additional fees dated 04.01.2001, the applicant has opted for neither to restrict, nor to pay additional fees. Consequently the IPE is limited to isolated nucleic acid fragments, that encode proteins that bind matrix attachment regions, wherein the protein is encoded by Seq. Id. No. 1/2 (claim 1(a) partially, and claims dependent thereof).

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/US00/09723

All other parts, namely those inventions listed under groups 2 and following, are excluded from IPE.

Re Item IV**Lack of unity of invention**

7. According to Rule 13 of the PCT regulations the requirement of unity of invention shall be fulfilled only when there is a technical relationship among those inventions involving one or more of the same or corresponding special technical features. The expression "special technical features" shall mean those technical features that define a contribution which each of the claimed inventions, considered as a whole, makes over the prior art.

In the present case this means that all the claimed molecules need to share on special technical feature which distinguishes the ensemble of the claimed molecules from the prior art, i.e. the molecule LeMFP1 known from D1. In particular, this means that all the claimed sequences need to contain at least one common particular structural feature which is responsible for a particular unexpected property, not found in LeMFP1.

- 7.1 However, no such feature is detectable. Consequently, in addition to the groups identified by the ISA, the IPEA is of the opinion that no unifying concept is linking above mentioned separate embodiments.

- 7.2 The separate inventions/groups of inventions are:

Isolated nucleic acid fragments, that encode proteins that bind matrix attachment regions, wherein the protein

group 1: is encoded by Seq. Id. No. 1/2 (claim 1(a) partially, and claims dependent thereof)

group 2: is encoded by Seq. Id. No. 3/4 (claim 1(a) partially, and claims dependent thereof)

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/US00/09723

group 3: is encoded by Seq. Id. No. 19/20 (claim 15(a) partially, and claims dependent thereof)

group 4: is encoded by Seq. Id. No. 21/22 (claim 29(a) partially, and claims dependent thereof)

group 5: is encoded by Seq. Id. No. 23/24 (claim 43(a) partially, and claims dependent thereof)

group 6 -:

As the fragments and variants of the above sequences (see item II-VII) fall apart into an uncountable number of separate molecules, it does not appear reasonable to invite to pay additional fees for said additional inventions. In any case, the applicant must be aware of that only said fragments and variants of the above 4 separate sequences will be covered by IPE which are unified, according to the criteria laid out in Rule 13, PCT, with at least one of the above 4 sequences.

- 7.3. As the applicant has neither restricted the claims nor has he paid additional fees, the IPE will be restricted to those parts of the application as specified under item 7.2. group 1, i.e. claims 1-14 (partially).

Re Item V

**Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
Novelty; Art 33(2), PCT**

8. The subject-matter proposed in claims 1-14 of the present application is considered formally novel.

The prior art protein LeMFP1 (D1), and sequences encoding it, falls under most of the structural definitions given in claim 1 and claims dependent thereof, such as being substantially similar to Seq. Id. No. 2 (see also Table 1 and 3 of the

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/US00/09723

application) and is also considered to fulfill any of the criteria set out for hybridization. However LeMFP1 does not encode a tobacco protein. As a consequence as the subject-matter of claims 1-14 is limited to MFP-like proteins from tobacco said claims are considered formally novel over D1.

Inventive Step; Art 33(3), PCT

9. The subject-matter proposed in claim 1-14 of the present application cannot be considered as involving an inventive step (Article 33(3) PCT) for the following reasons.

Claim 1-6 relates to a novel protein MFP-like protein termed NtMFP1 (defined by Seq. Id. No. 1/2). As however no further functions, as those which can be extrapolated from the structural relatedness to the known LeMFP1, are derivable from the present application, the disclosure of the present application appears to be limited to the identification of a further MFP-like protein. In order to provide an inventive contribution however, the mere fact of constituting a homolog to LeMFP1 is not sufficient. In order to establish inventive activity, the provision of a sequence must be justified by the technical purpose, i.e. by a hitherto unknown or unexpected technical effect, caused by those technical features which distinguish the claimed molecules from those of the prior art.

Due to the absence of any disclosed function or technical effect which goes beyond that already known for LeMFP1, the provision of any of the present sequence is considered to amount to nothing more than the provision of an equivalent which is considered to lack an inventive step, pursuant to Article 33(3) PCT.

Re Item VIII**Certain observations on the international application**

10. Claim 3 appears to be misworded, as it refers to polypeptides encoded by nucleic acids defined under claim 1 (c) and (e).

- 10.1 Claim 3 appears to be misworded, as it is directed to proteins encoded by the

INTERNATIONAL PRELIMINARY

International application No. PCT/US00/09723

EXAMINATION REPORT - SEPARATE SHEET

complement of the sequence depicted by Seq. Id. No. 1 (claim 1e). These putative proteins however are structurally and functionally completely unrelated to the protein of the invention and as such do neither contribute to the solution of the technical problem nor are they unified under the same concept.

10.2 The same applies in analogy to proteins encoded by nucleic acids of undefined length which hybridize to Seq. Id. No. 1 (claim 1(c)).

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference BC1003PCT	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/US 00/ 09723	International filing date (day/month/year) 12/04/2000	(Earliest) Priority Date (day/month/year) 12/04/1999
Applicant E.I. DU PONT DE NEMOURS AND COMPANY et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 5 sheets.



It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the language, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.



the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

- b. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of the sequence listing:



contained in the international application in written form.



filed together with the international application in computer readable form.



furnished subsequently to this Authority in written form.



furnished subsequently to this Authority in computer readable form.



the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.



the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☐ Certain claims were found unsearchable (See Box I).

3. ☒ Unity of invention is lacking (see Box II).

4. With regard to the title,

the text is approved as submitted by the applicant.



the text has been established by this Authority to read as follows:

5. With regard to the abstract,

the text is approved as submitted by the applicant.



the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the drawings to be published with the abstract is Figure No.

as suggested by the applicant.



because the applicant failed to suggest a figure.



because this figure better characterizes the invention.

4B

None of the figures.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/09723

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

1,2,3,4,7,8,11,12,13-20
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1,2,3,4,11,12,13-20 partially

Nucleic acid sequences derived from tobacco as characterized by SEQIDs 1,3,11,12,13,14,15 which encode a tobacco MFP1 protein as characterized by SEQIDs 2 + 4; the recombinant expression of the same in host cells and furthermore a method for altering the level of the MFP1 protein in a host cell by transforming the host cell with a chimeric gene as mentioned above; and a method to obtain related sequences by hybridisation or PCR using specific primers.

2. Claims: 1,2,3,4,11,12,13-20 partially; 7,8 completely

Nucleic acid sequence derived from soybean as characterized by SEQID 19 which encodes a soybean MFP1 protein as characterized by SEQIDs 20; the recombinant expression of the same in host cells and furthermore a method for altering the level of the MFP1 protein in a host cell by transforming the host cell with a chimeric gene as mentioned above; and a method to obtain related sequences by hybridisation or PCR using specific primers.

3. Claims: 1,2,3,4,11,12,13-20 partially; 5,6 completely

Nucleic acid sequence derived from corn as characterized by SEQID 21 which encodes a corn MFP1 protein as characterized by SEQIDs 22; the recombinant expression of the same in host cells and furthermore a method for altering the level of the MFP1 protein in a host cell by transforming the host cell with a chimeric gene as mentioned above; and a method to obtain related sequences by hybridisation or PCR using specific primers.

4. Claims: 1,2,3,4,11,12,13-20 partially; 9,10 completely

Nucleic acid sequence derived from rice as characterized by SEQID 23 which encodes a rice MFP1 protein as characterized by SEQIDs 24; the recombinant expression of the same in host cells and furthermore a method for altering the level of the MFP1 protein in a host cell by transforming the host cell with a chimeric gene as mentioned above; and a method to obtain related sequences by hybridisation or PCR using specific primers.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/09723

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C07K14/415 C12N5/10 C12Q1/68 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, EPO-Internal, STRAND, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MEIER IRIS ET AL: "MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA." PLANT CELL, vol. 8, no. 11, 1996, pages 2105-2115, XP002139228' ISSN: 1040-4651 cited in the application the whole document	1,3,13, 14,16
A	MEIER, I., ET AL.: "MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA" EMBL SEQUENCE DATA LIBRARY, 8 January 1997 (1997-01-08), XP002143996' heidelberg, germany accession no. Y07861 --- -/-	

☒ Further documents are listed in the continuation of box C.☐ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

5 October 2000

Date of mailing of the international search report

18. 10. 00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/09723

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	HARDER PATRICIA A ET AL: "Conservation of matrix attachment region-binding filament-like protein 1 among higher plants." PLANT PHYSIOLOGY (ROCKVILLE), vol. 122, no. 1, January 2000 (2000-01), pages 225-234, XP002143997 , ISSN: 0032-0889 the whole document ----	1-4
P,X	DATABASE EMBL SEQUENCE DATABASE 'Online! 28 July 1999 (1999-07-28) SHOEMAKER, R., ET AL. : "public soybean EST project " XP002149331\, accession no. AI901273 -----	1
T	GINDULLIS FRANK ET AL: "Matrix attachment region binding protein MFP1 is localized in discrete domains at the nuclear envelope." PLANT CELL, vol. 11, no. 6, June 1999 (1999-06), pages 1117-1128, XP002143998 . ISSN: 1040-4651 the whole document -----	

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 October 2000 (19.10.2000)

PCT

(10) International Publication Number
WO 00/61615 A3

(51) International Patent Classification⁷: C12N 15/82,
C07K 14/415, C12N 5/10, C12Q 1/68, A01H 5/00

A. [US/US]; 712 West 34th Street, Wilmington, DE 19802
(US). MEIER, Iris [DE/US]; 4467 Masters Drive, Colum-
bus, OH 43220 (US).

(21) International Application Number: PCT/US00/09723

(22) International Filing Date: 12 April 2000 (12.04.2000)

(74) Agent: FELTHAM, S., Neil; E.I. Du Pont de Nemours
and Company, Legal Patent Records Center, 1007 Market
Street, Wilmington, DE 19898 (US).

(25) Filing Language: English

(26) Publication Language: English

(81) Designated States (*national*): BR, US.

(30) Priority Data:
60/128,900 12 April 1999 (12.04.1999) US

(84) Designated States (*regional*): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).

(71) Applicant (*for all designated States except US*): E.I. DU
PONT DE NEMOURS AND COMPANY [US/US]; 1007
Market Street, Wilmington, DE 19898 (US).

Published:
— With international search report.

(72) Inventors; and

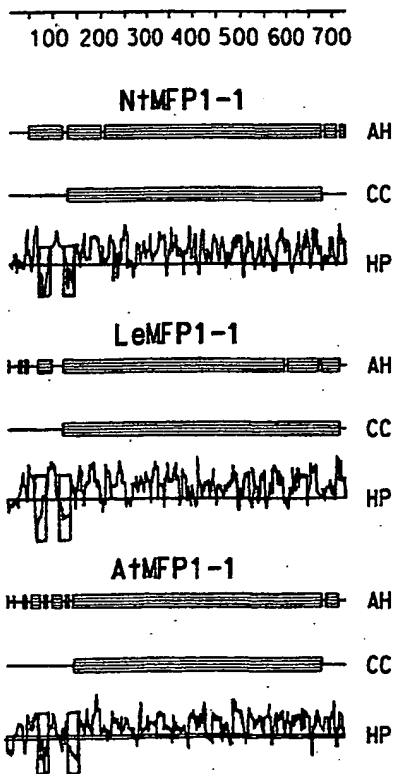
(88) Date of publication of the international search report:
12 April 2001

(75) Inventors/Applicants (*for US only*): HARDER, Patricia,

[Continued on next page]

(54) Title: HOMOLOGS OF MAR-BINDING FILAMENT-LIKE PROTEIN 1 (MFP1)

(57) Abstract: This invention pertains to nucleic acid fragments en-
coding proteins that are homologs to the MAR binding filament-like
protein 1 (MFP1) from tomato. More specifically, this invention per-
tains to two tobacco MFP1 genes and MFP1 homologs from corn,
soybean, and rice.



RECEIVED

APR 23 2001

PATENT RECORDS
CENTER

WO 00/61615 A3



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLEHOMOLOGS OF MAR-BINDING
FILAMENT-LIKE PROTEIN 1 (MFP1)

This application claims the benefit of U.S. Provisional Application
5 No. 60/128,900, filed April 12, 1999.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. This invention
pertains to nucleic acid fragments encoding proteins that are homologs to the
MAR-binding filament-like protein 1 (MFP1) from tomato. More specifically,
10 this invention pertains to two tobacco MFP1 genes and MFP1 homologs from
corn, soybean and rice.

BACKGROUND OF THE INVENTION

The nuclear matrix hypothesis proposes a structural framework for the
eukaryotic nucleus that is similar to the cytoskeleton. To date, its best
15 characterized component is the lamina, a filamentous protein network that lines
the inner membrane of the nuclear envelope. Major components of the lamina
include a group of intermediate-filament (IF) proteins, collectively known as
nuclear lamins, that are classified as type A, B, and C (McKeon *et al.*, *Nature*
319:463-468 (1986)). Lamin B is attached to the inner nuclear membrane via a
20 C-terminal C15 farnesyl group (Schafer *et al.*, *Annu. Rev. Genet.* 30:209-237
(1992)), whereas lamins A and C bind to lamin B. Other integral membrane
proteins interact with lamin B and most likely stabilize the membrane attachment
of lamins (Furukawa *et al.*, *EMBO J.* 14:1626-1636 (1995)). Recent studies have
also demonstrated the ability of lamins A and B to bind DNA, suggesting a role
25 for mammalian lamins in anchoring chromatin to the nuclear envelope. The
interaction between nuclear envelope, lamina, and chromatin is considered to be
of fundamental importance for higher order chromosome organization, as well as
the assembly and disassembly of the nuclear envelope during mitosis (Furukawa
et al., *EMBO J.* 14:1626-1636 (1995)).

30 The nuclear matrix is a second structural skeleton that has been
biochemically defined as the insoluble component that remains after treatment of
isolated nuclei with DNase I and extraction of proteins with high-salt solutions
(Berezney *et al.*, *Biochem. Biophys. Res. Comm.* 60:1410-1417 (1974)) or the
chaotropic agent lithium diiodosalicylate (Mirkowitch *et al.*, *Cell* 39:223-232
35 (1984)). Chromatin binds to the nuclear matrix via matrix attachment regions
(MARs) in the DNA. MARs are generally AT-rich DNA sequences that are
several hundred base pairs long and localized to noncoding regions of the DNA,
but often flanking genes (Gasser *et al.*, *Trends Genet.* 3:16-22 (1987)). However,

there is no consensus sequence known for MARs. The significance of structural characteristics for MARs such as DNA bending and a narrow minor groove due to oligo(dA) tracts has been previously proposed. MARs have been shown to increase transcriptional activity of a linked gene and to confer position-independent, copy-number dependent expression in stably transfected cells (Phi-Wan et al., *EMBO J.* 7:655-664 (1988)).

A small number of MAR binding proteins have been identified from animal nuclei, and they are considered to be components of the nuclear matrix (von Kries et al., *Cell* 64:123-135 (1991); Dickinson et al., *Cell* 70:631-645 (1992); Romig et al., *EMBO J.* 11:3431-3440 (1992); Tsutsui et al., *J. Biol. Chem.* 268:12886-12894 (1993); Renz et al., *Nucleic Acids Res.* 24:843-849 (1996); U.S. 5,652,340). In addition, it has been shown that lamins specifically bind to MARs (Luderus et al., *Mol. Cell. Biol.* 14:6297-6305 (1994)). The specific interaction between DNA and the nuclear matrix/nuclear lamina is most likely an important mechanism for long-range gene regulation and higher order chromatin organization (Gasser et al., *Trends Genet.* 3:16-22 (1987)).

Most investigations into structural components of the nucleus have focused on proteins in vertebrates and *Drosophila*, but even in these organisms, our knowledge about the molecular constituents of the nuclear matrix is sparse. Significantly less information is available for other eukaryotes, and in particular for plants. Proteins that are immunologically related to animal IF proteins and lamins have been detected in pea and carrot nuclei (Beven et al., *J. Cell Sci.* (1991) 98 (3), 293-30; McNulty et al., *J. Cell Sci.* 103:407-414 (1992)). Plant nuclear matrix preparations that bind to animal MARs have been reported, suggesting that proteins with similar DNA binding specificities exist in plants as well (Hall et al., *Proc. Natl. Acad. Sci. USA* 88:9320-9324 (1991)).

Effects of MARs on gene expression in plants have been reported, but have been quite variable. In some experimental systems, no reduction of variability but an increase in expression level has been reported (Breyne et al., *Plant Cell* 4:463-471 (1992); Allen et al., *Plant Cell* 5:603-613 (1993); Allen et al., *Plant Cell* 8:899-913 (1996); U.S. 5,773,689). Other authors have found no significant increase in expression level, but a reduction of variability (van der Geest et al., *Plant J.* 6:413-423 (1994); Mlynarova et al., *Plant Cell* 6:417-426 (1994)). It is not clear what causes these observed differences, but they will most probably be due to the fact that MARs establish different molecular interactions, which might either depend on the features of the MAR itself or on the specific molecular environment of the transformed cell/tissue. The routine use of MARs for strategies to improve transgene expression will greatly depend on the

characterization of the proteins involved in DNA-nuclear matrix attachment and the factors responsible for the observed increase in gene expression.

Currently, no sequence information is available for plant lamin-like proteins. However, the cloning of the cDNA for a plant MAR-binding protein, MFP1, from tomato has been reported (Meier et al., *Plant Cell* 8:2105-2115 (1996)). MFP1 has structural features of a filament-like protein and it preferentially binds to MAR DNA sequences from both plants and animals. In contrast to other known MAR binding proteins, MFP1 contains a hydrophobic N-terminal amino acid sequence that might function as a membrane-spanning domain. MFP1, therefore, has features of a novel anchor protein that most likely connects chromatin via MAR DNA with the nuclear envelope and nuclear filament proteins.

In order to routinely use the attachment of transgenes to the nuclear matrix improve gene expression, it will be necessary to further characterize the elements involved in this process and to better understand the underlying mechanisms. Thus, a need exists to identify and characterize additional nuclear matrix proteins. The present invention presents MFP1-like proteins from other plant species. Furthermore, the present invention shows that a single, immunologically related protein of comparable size is present in a variety of higher-plant species, including important crop plants. This invention pertains to the isolation of cDNAs corresponding to two tobacco MFP1 genes and the characterization of the MFP1 gene family in tobacco. The invention also pertains to the identification and partial characterization of EST sequences from corn, soybean and rice encoding MFP1 proteins from these crop species.

SUMMARY OF THE INVENTION

The present invention provides an isolated nucleic acid fragment encoding a plant MFP1 protein selected from the group consisting of: (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24; (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24; (c) an isolated nucleic acid molecule that hybridizes with a nucleic acid sequence of (a) or (b) under the following hybridization conditions: 5 x Denhards, 5 x SSPE, 5% SDS, 20 µg/mL salmon sperm DNA at 55 °C; (d) an isolated nucleic acid molecule that hybridizes with a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3,

SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21 and SEQ ID NO:23 under the following hybridization conditions: 5 x Denhards, 5 x SSPE, 5% SDS, 20 µg/mL salmon sperm DNA at 55 °C; and (e) an isolated nucleic acid fragment that is

5 complementary to (a), (b), (c) or (d).

Additionally the invention provides a nucleic acid fragment, isolated from corn, encoding an MFP1 polypeptide, the polypeptide having at least 40% identity to SEQ ID NO:17, over a length of about 672 amino acids as compared by the Jotun-Hein algorithm.

10 Similarly the invention provides a nucleic acid fragment, isolated from soybean, encoding an MFP1 polypeptide, the polypeptide having at least 46% identity to SEQ ID NO:17 over a length of 388 amino acids as compared by the Jotun-Hein algorithm.

In another embodiment the invention provide a nucleic acid fragment, isolated, from rice, encoding an MFP1 polypeptide, the polypeptide having at least 39% identity to SEQ ID NO:17 over a length of 107 amino acids as compared by the Jotun-Hein algorithm.

15 In an alternate embodiment the invention provides an isolated nucleic acid fragment encoding a plant MFP1 polypeptide, the peptide having at least 77% identity to SEQ ID NO:17.

20 The invention further provides polypeptides encoded by the isolated nucleic acid fragments of the present invention.

In another embodiment the invention provides a chimeric gene comprising the isolated nucleic acid fragment of the present invention operably linked to suitable regulatory sequences.

25 The invention additionally provides a method of altering the level of expression of a plant MFP1 protein in a host cell comprising: (a) transforming a host cell with the chimeric gene of the present invention and; (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene resulting in production of altered levels of a plant MFP1 protein in the transformed host cell relative to expression levels of an untransformed host cell.

30 The invention additionally provides transformed host cells comprising the chimeric genes of the present invention.

35 In an alternate embodiment the invention provides methods of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a plant MFP1 protein using portions of the present nucleic acid sequences as hybridization probes or as primers.

**BRIEF DESCRIPTION OF THE DRAWINGS,
AND SEQUENCE DESCRIPTIONS**

Figure 1 shows a schematic representation of the subfragments E-196 and H-207 that were expressed in *Escherichia coli*.

5 Figure 2A is a gel showing the immunological identification of MFP1-like proteins in different plant species using the aR50 antibody raised against a Le MFP1 polypeptide.

Figure 2B is a gel showing the immunological identification of MFP1-like proteins in different plant species using the a288 antibody raised against a Le
10 MFP1 polypeptide.

Figure 3 shows the schematic structure of the partial cDNAs isolated from a *Nicotiana tabacum* lambda ZAP cDNA library.

Figure 4A shows the percent identical amino acids in pairwise comparisons of the four MFP1 proteins.

15 Figure 4B shows the hydrophilicity and secondary structure analysis of LeMFP1, NtMFP1-1 and AtMFP1.

Figure 5 shows the genomic organization of tobacco MFP1.

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions which form part of this
20 application.

The following sequence descriptions and sequence listings attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825. The Sequence Descriptions contain the one letter code for nucleotide sequence
25 characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219(2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

30 SEQ ID NO:1 is the nucleotide sequence for NtMFP1-1.

SEQ ID NO:2 is the deduced amino acid sequence for NtMFP1-1, encoded by SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence for NtMFP1-2.

35 SEQ ID NO:4 is the deduced amino acid sequence for NtMFP1-2, encoded by SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence which codes for E-196 polypeptide fragment isolated from tomato.

SEQ ID NO:6 is the deduced amino acid sequence for E-196 polypeptide fragment isolated from tomato, encoded by SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence which codes for H-207 polypeptide fragment isolated from tomato.

5 SEQ ID NO:8 is the amino acid sequence for H-207 polypeptide fragment isolated from tomato, encoded by SEQ ID NO:8.

SEQ ID NO:9 is the nucleotide sequence for the p7-2 fragment isolated from tomato.

10 SEQ ID NO:10 is the nucleotide sequence for the p1-3 fragment isolated from tomato.

SEQ ID NO:11 is the nucleotide sequence for the T6 fragment isolated from tobacco.

SEQ ID NO:12 is the nucleotide sequence for the T1 fragment isolated from tobacco.

15 SEQ ID NO:13 is the nucleotide sequence for the T2 fragment isolated from tobacco.

SEQ ID NO:14 is the nucleotide sequence for the T3 fragment isolated from tobacco.

20 SEQ ID NO:15 is the nucleotide sequence for the PCR1 fragment isolated from tobacco.

SEQ ID NO:16 is the nucleotide sequence for LeMFP1.

SEQ ID NO:17 is the deduced amino acid sequence for LeMFP1, encoded by SEQ ID NO:16.

SEQ ID NO:18 is the nucleotide sequence used as a Southern probe.

25 SEQ ID NO:19 is the nucleotide sequence comprising the cDNA insert in clone src3c.pk004.m1 encoding a soybean MFP1 (GmMFP1).

SEQ ID NO:20 is the deduced amino acid sequence of the nucleotide sequence comprising the cDNA insert in clone src3c.pk004.m1.

30 SEQ ID NO:21 is the nucleotide sequence comprising the cDNA insert in clone p0118.chsab48r encoding a corn MFP1.

SEQ ID NO:22 is the deduced amino acid sequence of the nucleotide sequence comprising the cDNA insert in clone p0118.chsab48r.

SEQ ID NO:23 is the nucleotide sequence comprising the cDNA insert in clone rca1n.pk022.a11 encoding a rice MFP1.

35 SEQ ID NO:24 is the deduced amino acid sequence of the nucleotide sequence comprising the cDNA insert in clone rca1n.pk022.a11.

SEQ ID NO:25 is the nucleotide sequence for PCR primer designed from T3 fragment isolated from tobacco.

SEQ ID NO:26 is the nucleotide sequence for PCR primer designed from T1 fragment isolated from tobacco.

DETAILED DESCRIPTION OF THE INVENTION

The present invention reports the isolation and characterization of cDNAs corresponding to two tobacco MFP1 genes and the isolation and identification of MFP1 EST homologs from corn, soybean and rice. No homologs of MFP1 from tobacco have been described previously. The level of expression of the genes described here can be altered in the plant by methods of cosuppression and overexpression. As they are previously undescribed genes involved in a fundamental cellular mechanism, this can lead to novel developmental phenotypes that might be beneficial for crop growth and development. In addition, if the reduction in expression of one of the genes leads to a growth or developmental defect in the plant, this gene can be used as a novel herbicide target. All isolated proteins can be used as tools to study the plant nuclear matrix, of which no components have been isolated at the molecular level. This can lead to the identification of additional proteins, that can be used as described above. Any related EST sequences can be directly used for the above described applications in crop plants. All of these sequences can be directly used to broaden our understanding of the mechanisms of MAR-matrix interactions and the molecular basis for the described effects on gene expression.

The following definitions are provided for the full understanding of terms and abbreviations used in this specification.

"Polymerase chain reaction" is abbreviated PCR.

"Expressed sequence tag" is abbreviated EST.

"Open reading frame" is abbreviated ORF.

"SDS polyacrylamide gel electrophoresis" is abbreviated SDS-PAGE.

"Amino acid" is abbreviated AA.

"Plaque-forming units" is abbreviated pfus.

" α -Helical" is abbreviated AH.

"Coiled-coil" is abbreviated CC and refers to an amphiphilic α -helical protein structure.

"Hydrophilicity plot" is abbreviated HP.

"Matrix attachment region" is abbreviated MAR. MARs are also known as matrix-associated regions or scaffold-associated (or attachment) regions.

The term "MFP" is an abbreviation for MAR-binding filament-like protein. "MFP1" refers to the MAR-binding filament-like protein having similar characteristics to the protein isolated from tomato as described in Meier et al., *Plant Cell* 8:2105-2115 (1996). "LeMFP1" is the abbreviation for the specific

MFP1 protein isolated from tomato, as set forth in SEQ ID NO:17. "NtMFP1-1" and "NtMFP1-2" are the abbreviations for the first and second MFP1 proteins isolated from tobacco, as set forth in SEQ ID NO:2 and 4 respectively.

5 "GmMFP1" is the abbreviation for the MFP1 protein isolated from soybean, as set forth in SEQ ID NO:20. "ZmMFP1" is the abbreviation for the MFP1 protein isolated from corn, as set forth in SEQ ID NO:22. "OsMFP1" is the abbreviation for the MFP1 protein isolated from rice, as set forth in SEQ ID NO:24. "AtMFP1" is the abbreviation for the MFP1 protein isolated from Arabidopsis, released on (<http://genomewww.standard.edu/Arabidopsis/>).

10 The terms "isolated nucleic acid fragment" or "isolated nucleic acid molecule" refer to a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment or an isolated nucleic acid molecule in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, or synthetic DNA.

The terms "host cell" and "host organism" refer to a cell capable of receiving foreign or heterologous genes and expressing those genes to produce an active gene product. Suitable host cells include microorganisms such as bacteria and fungi, as well as plant cells.

20 The term "fragment" refers to a DNA or amino acid sequence comprising a subsequence of the nucleic acid sequence or protein of the present invention. However, an active fragment of the present invention comprises a sufficient portion of the protein to maintain activity.

The term "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases result in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases do not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the present invention such as deletion or insertion of one or more nucleotide bases that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

35 A "substantial portion" refers to an amino acid or nucleotide sequence which comprises enough of the amino acid sequence of a polypeptide or the

nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al., *J. Mol. Biol.* 215:403-410 (1993); see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides (generally 12 bases or longer) may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The present specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for the purpose known to those skilled in the art. Accordingly, the present invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% identity with the gene to be suppressed. Moreover, alterations in a gene that result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of

the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their ability to
5 hybridize, under stringent conditions (0.1 x SSC, 0.1% SDS, 65 °C), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the present invention are those nucleic acid fragments whose DNA sequences are 80% identical to the DNA sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are 90% identical to the
10 DNA sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are 95% identical to the DNA sequence of the nucleic acid fragments reported herein.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid
15 sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR,
20 Inc. 1228 S. Park St. Madison, WI 53715 USA). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default
25 vales" will mean any set of values or parameters which originally load with the software when first initialized.

The term "percent identity" is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be,
30 as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993);
35 Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but

5 are not limited to, the GCG Pileup program found in the GCG program package, using the Needleman and Wunsch algorithm with their standard default values of gap creation penalty=12 and gap extension penalty=4 (Devereux *et al.*, *Nucleic Acids Res.* 12:387-395 (1984)), BLASTP, BLASTN, and FASTA (Pearson *et al.*, *Proc. Natl. Acad. Sci. USA* 85:2444-2448 (1988)). The BLASTX program is

10 publicly available from NCBI and other sources (BLAST Manual, Altschul *et al.*, Natl. Cent. Biotechnol. Inf., Natl. Library Med. (NCBI NLM) NIH, Bethesda, Md. 20894; Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990); Altschul *et al.*, *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*, *Nucleic Acids Res.* 25:3389-3402 (1997)). The method to determine percent

15 identity preferred in the present invention is by the method of DNASTAR protein alignment protocol using the Jotun-Hein algorithm (Hein *et al.*, *Methods Enzymol.* 183:626-645 (1990)). Default parameters used for the Jotun-Hein method for alignments are: for multiple alignments, gap penalty=11, gap length penalty=3; for pairwise alignments ktuple=2. As an illustration, for a polynucleotide having a

20 nucleotide sequence with at least 95% identity to a reference nucleotide sequence, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95%

25 identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide

30 sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, for a polypeptide having an amino acid sequence having at least 95% "identity" to a reference amino acid sequence, it is intended that the amino acid sequence of the

35 polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up

to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

“Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the present invention relates to any nucleic acid fragment that encodes all or a substantial portion of present MFP1 proteins as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell to use nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

The term “complementary” is used to describe the relationship between nucleotide bases that are hybridizable to one another. Hence with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

A nucleic acid molecule is “hybridizable” to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the “stringency” of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a T_m of 55°, can be used, e.g., 5X SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5X SSC, 0.5% SDS. Moderate stringency hybridization conditions correspond to a higher T_m , e.g., 40% formamide, with 5X or 6X SSC.

Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of

5 complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater

10 than 100 nucleotides in length, equations for calculating T_m have been derived (see Sambrook et al., *supra*, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., *supra*, 11.7-11.8). In one embodiment the length for a hybridizable nucleic

15 acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such

20 as length of the probe.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. "Chemically

25 synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on

30 optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determining preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

35 "Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene"

refers to any gene, not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but which is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a DNA sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (*Biochemistry of Plants* 15:1-82 (1989)). It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner et al., *Mol. Biotech.* 3:225 (1995)).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (*Plant Cell* 1:671-680 (1989)).

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA.

"Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

The term "operably-linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably-linked with a coding sequence when it affects the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

The term "expression" refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the

invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds
5 levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. 5,231,020).

"Altered levels" refers to the production of gene product(s) in organisms
10 in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of
15 translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made.
20 "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can
25 further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the
30 genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al., *Meth. Enzymol.* 143:277 (1987)) and particle-accelerated or "gene gun" transformation technology
35 (Klein et al., *Nature, London* 327:70-73 (1987); U.S. 4,945,050).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch,

E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Novel MFP1-binding proteins, have been isolated from tobacco, corn, soybean and rice. Comparison of their random cDNA sequences to the GenBank
5 database using the BLAST and DNASTAR algorithms, well known to those skilled in the art, revealed that these proteins have no significant homologies to other known proteins, other than MFP1 proteins. The nucleotide sequences of the present MFP1 cDNA are provided in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ
10 ID NO:19, SEQ ID NO:21 and SEQ ID NO:23. Other MFP1 genes and proteins from other plants can now be identified by comparison of random cDNA sequences to the present MFP1 sequences provided herein.

Comparison of the instant MFP1 base deduced amino acid sequences to the only published sequence of this kind (LeMFP1, Meier *et al.*, *Plant Cell*
15 8:2105-2115 (1996); SEQ ID NO:17 and 18) show a variation of homology of about 39% identity (rice, SEQ ID NO:24) over a length of 107 amino acids to about 77% identity for tobacco (SEQ ID NO:2 and 4) as compared by the Jotun-Hein alignment algorithm (Hein *et al.*, *Methods Enzymol.* 183:626-645 (1990)).

20 Accordingly preferred polypeptides of the instant invention are those plant proteins which are at least 77% identical to the amino acid sequence as set forth in SEQ ID 17. More preferred amino acid fragments are at least about 80%-90% identical to the sequences herein. Most preferred are nucleic acid fragments that are at least 95% identical to the amino acid fragments reported herein. Similarly,
25 preferred nucleic acid sequences are those encoding MFP1 binding proteins and which are at least 80% identical to the nucleic acid sequences of reported herein. More preferred nucleic acid fragments are at least 90% identical to the sequences herein. Most preferred are nucleic acid fragments that are at least 95% identical to the nucleic acid fragments reported herein.

30 Similarly preferred polypeptides are those isolated from corn which are at least 40% identical to the polypeptide of SEQ ID NO:17 over a length of about 672 amino acids as compare by the Jotun-Hein alignment algorithm (Hein *et al.*, *supra*). Other preferred polypeptides are those isolated from rice which are at least 39% identical to the polypeptide of SEQ ID NO:17 over a length of about
35 107 amino acids as compare by the Jotun-Hein alignment algorithm (Hein *et al.*, *supra*). Additionally preferred polypeptides are those isolated from soybean which are at least 46% identical to the polypeptide of SEQ ID NO:17 over a

length of about 388 amino acids as compare by the Jotun-Hein alignment algorithm (Hein et al., *supra*).

The nucleic acid fragments of the present invention may be used to isolate cDNAs and genes encoding a homologous MFP1 proteins from the same or other plant species. Isolating homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction (PCR) or ligase chain reaction).

For example, other MFP1 genes, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the present nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant using methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the present MFP1 sequences can be designed and synthesized by methods known in the art (Maniatis, *supra*). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primers, DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of or full-length of the present sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the present nucleic acid fragment may be used in PCR protocols to amplify longer nucleic acid fragments encoding homologous MFP1 genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the present nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant MFP1.

Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *Proc. Natl. Acad. Sci., USA* 85:8998 (1988)) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the present sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., *Proc. Natl. Acad. Sci., USA* 86:5673 (1989); Loh

et al., *Science* 243:217 (1989)). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman et al., *Techniques* 1:165 (1989)).

Finally, availability of the present nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the present amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner et al., *Adv. Immunol.* 36:1 (1984); Maniatis, *supra*).

The nucleic acid fragments of the present invention may also be used to create transgenic plants in which the present MFP1 protein is present at higher or lower levels than normal. Alternatively, in some applications, it might be desirable to express the present MFP1 protein in specific plant tissues and/or cell types, or during developmental stages in which they would normally not be encountered.

Overexpression of the present MFP1 may be accomplished by first constructing a chimeric gene in which the MFP1 coding region is operably-linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals must also be provided. The present chimeric genes may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the present chimeric genes can then be constructed. The choice of a plasmid vector depends upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., *EMBO J.* 4:2411-2418 (1985); De Almeida et al., *Mol. Gen. Genetics* 218:78-86 (1989)), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the MFP1 protein to different cellular compartments or to facilitate their secretion from the cell. The chimeric genes described above may be further modified by the addition of appropriate intracellular or extracellular targeting sequence to their coding regions. These include chloroplast transit peptides (Keegstra et al., *Cell* 56:247-253 (1989), signal sequences that direct proteins to the endoplasmic reticulum (Chrispeels et al., *Ann. Rev. Plant Phys. Plant Mol.* 42:21-53 (1991), and nuclear localization signal (Raikhel et al., *Plant Phys.* 100:1627-1632 (1992). While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of the MFP1 genes in plants for some applications. In order to accomplish this, chimeric genes designed for antisense or co-suppression of MFP1 can be constructed by linking the genes or gene fragments encoding parts of these enzymes to plant promoter sequences. Thus, chimeric genes designed to express antisense RNA for all or part of MFP1 can be constructed by linking the MFP1 genes or gene fragments in reverse orientation to plant promoter sequences. The co-suppression or antisense chimeric gene constructs could be introduced into plants via well known transformation protocols wherein expression of the corresponding endogenous genes are reduced or eliminated.

The present MFP1 proteins may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the proteins by methods well known to those skilled in the art. The antibodies would be useful for detecting the present MFP1 protein *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the present MFP1 protein are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the present MFP1. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the present MFP1 protein.

Microbial host cells suitable for the expression of the present MFP1 proteins include any cell capable of expression of the chimeric genes encoding these proteins. Such cells will include both bacteria and fungi including, for example, the yeasts (e.g., *Aspergillus*, *Saccharomyces*, *Pichia*, *Candida*, and *Hansenula*), members of the genus *Bacillus* as well as the enteric bacteria (e.g., *Escherichia*, *Salmonella*, and *Shigella*). Methods for the transformation of such hosts and the expression of foreign proteins are well known in the art and

examples of suitable protocols may be found in Manual of Methods for General Bacteriology (Gerhardt et al., eds., American Society for Microbiology, Washington, DC. (1994)) or in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Brock, T. D., Sinauer Associates, Inc., Sunderland, MA (1989)).

Vectors or cassettes useful for transforming suitable microbial host cells are well known in the art. Typically the vector or cassette contains sequences directing transcription and translation of the relevant gene, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell, although such control regions need not be derived from the genes native to the specific species chosen as a production host.

Initiation control regions or promoters useful to drive expression of the genes encoding the MFP1 proteins in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including but not limited to CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI (useful for expression in *Saccharomyces*); AOX1 (useful for expression in *Pichia*); and lac, trp, lP_L , lP_R , T7, tac, and trc (useful for expression in *E. coli*). Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

Additionally, the present MFP1 proteins can be used as targets to facilitate the design and/or identification of inhibitors of MFP1 that may be useful as herbicides or fungicides. This could be achieved either through the rational design and synthesis of potent functional inhibitors that result from structural and/or mechanistic information that is derived from the purified present plant proteins, or through random *in vitro* screening of chemical libraries. It is anticipated that significant *in vivo* inhibition of any of the MFP1 proteins described herein may severely cripple cellular metabolism and likely result in plant (or fungal) death.

All or a portion of the nucleic acid fragments of the present invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to expression of the present MFP1. Such information may be useful in plant breeding in order to develop lines with

desired phenotypes. For example, the present nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis, *supra*) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the present invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al., *Genomics* 1:174-181 (1987)) in order to construct a genetic map. In addition, the nucleic acid fragments of the present invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the present nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al., *Am. J. Hum. Genet.* 32:314-331 (1980)).

The production and use of plant gene-derived probes for use in genetic mapping is described by Bernatzky et al. (*Plant Mol. Biol. Reporter* 4:37-41 (1986)). Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the present nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al., Nonmammalian Genomic Analysis: A Practical Guide, pp. 319-346, Academic Press (1996), and references cited therein).

In another embodiment, nucleic acid probes derived from the present nucleic acid sequence may be used in direct fluorescence *in situ* hybridization (FISH) mapping. Although current methods of FISH mapping favor use of large clones (several to several hundred kb), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the present nucleic acid sequences. Examples include allele-specific amplification (Kazazian et al., *J. Lab. Clin. Med.* 114:95-96 (1989)), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al., *Genomics* 16:325-332 (1993)), allele-specific ligation (Landegren et al., *Science* 241:1077-1080 (1988)), nucleotide extension reactions (Sokolov et al., *Nucleic Acid Res.* 18:3671 (1990)), Radiation Hybrid Mapping (Walter et al., *Nature Genetics* 7:22-28 (1997)) and Happy Mapping (Dear et al., *Nucleic Acid Res.* 17:6795-6807 (1989)). For these methods, the sequence of a nucleic acid

fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods using PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the present nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function-mutant phenotypes may be identified for the present cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger et al., *Proc. Natl. Acad. Sci. USA* 86:9402 (1989); Koes et al., *Proc. Natl. Acad. Sci. USA* 92:8149 (1995); Bensen et al., *Plant Cell* 7:75 (1995)). The latter approach may be accomplished in two ways. First, short segments of the present nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the MFP1 protein. Alternatively, the present nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a MFP1 protein can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the MFP1 gene product.

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usage and conditions.

35 EXAMPLES

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook et al. (1989), J., Fritsch,

E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989 (hereinafter "Maniatis"); and by T. J. Silhavy, M. L. Bannan, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring, N.Y. (1984) and by
5 Ausubel et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987).

Nucleotide and amino acid percent identity and similarity comparisons were made using the DNASTAR suite of programs, applying default parameters unless indicated otherwise.

10 The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter, "mL" means milliliters, "L" means liters, "mM" means millimolar, "M" means molar, "mmol" means millimole(s).

Plant material and growth conditions:

15 Tobacco, tomato, soybean, rice, corn, wheat and *Arabidopsis thaliana* were grown in soil in a growth chamber with a 12 h 24 °C light cycle followed by a 12 h, 20 °C dark cycle.

EXAMPLE 1

Isolation Of Total Protein

20 Total protein extracts were prepared from leaf tissues. 100 mg aliquots of tissue were ground to a fine powder with mortar and pestle in liquid nitrogen, resuspended in 0.5 mL extraction buffer (62.5 mM Tris-Cl, pH 6.8, 20% glycerol, 4% SDS, and 1.4 M β-mercaptoethanol) and incubated at 70 °C for 10 min. The debris was removed by centrifugation at 15,000 rpm for 10 min at 4 °C. The
25 supernatants were removed to new tubes, frozen in liquid nitrogen, and stored at -80 °C.

EXAMPLE 2

Protein Expression, Purification And Antibody Production

pRSETC-MFP1-EcoRI (containing the coiled-coil domain) and
30 pRSETA-MFP1-HincII (containing the DNA binding domain), the expression vectors for E-196 (SEQ ID NO:6) coded by SEQ ID NO:5 and H-207 (SEQ ID NO:8) coded by SEQ ID NO:7 fragments, respectively, have been described previously (Meier *et al.*, *Plant Cell* 8:2105-2115 (1996)). Figure 1 shows a representation of the subfragments E-196 and H-207 that were expressed in
35 *Escherichia coli*. Filled bars indicate α-helical regions, open bars indicate hydrophobic domains. The shaded box marks the DNA-binding domain. Numbers indicate the position of the first and last amino acid of each subfragment. Expression of recombinant fusion proteins containing an N-terminal 6-histidine

tag fused to the protein subfragments E-196 (SEQ ID NO:6) and H-207 (SEQ ID NO:8) was induced by isopropyl-D-thiogalactoside in *Escherichia coli* BL21 cells according to the Qiagen protein expression manual (Qiagen, Chatsworth, CA). The amount of fusion protein present in the different total *E. coli* protein extracts was determined by immunoblotting (Maniatis) with a monoclonal antibody directed against the T7 tag (Novagen, Madison, Wisconsin). The expressed proteins were purified by nickel-affinity chromatography (Qiagen, Chatsworth, CA), followed by SDS PAGE. The bands corresponding to the fusion proteins (~1 mg each) were excised from the gel, ground and used to raise two rabbit antisera (a288 against E-196 (SEQ ID NO:6) and aR50 against H-207 (SEQ ID NO:8)). Polyclonal antibodies were produced in rabbits by Eurogentech, Belgium, using the company's standard immunization protocols. The a288 antibody has been described previously (Meier et al., *Plant Cell* 8:2105-2115 (1996)).

EXAMPLE 3

Immuno-detection Of MFP1 Related Proteins In A Variety Of Higher Plant Species

A 1:3000 dilution of a288 or aR50 antiserum, and a 1:5000 dilution of horseradish peroxidase-coupled anti-rabbit secondary antibody (Amersham, Buckinghamshire, England) were used to perform immunoblot analyses (Maniatis). Enhanced chemiluminescence detection was performed using an ECL detection kit as described by the manufacturer (Amersham Buckinghamshire, England).

a288 and aR50 polyclonal antibodies were then used to detect proteins with antigenic similarity to MFP1 in other plant species (Figure 2). Total protein extracts were prepared from mature leaf tissues of tomato (*Lycopersicon esculentum* L.), tobacco (*Nicotiana tabacum* L.), *Arabidopsis thaliana*, soybean (*Glycine max* L.), rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), and corn (*Zea mays* L.) as described above. Equal amounts of total protein, as determined by Coomassie Brilliant Blue staining of a replica protein gel, were probed in immunoblot experiments with aR50 (Figure 2A) and a288 (Figure 2B) polyclonal antibodies. The arrow indicates the position of the MFP1-like proteins of approximately equal size in both panels. The position of molecular weight markers is indicated. aR50 antibody detects a single protein of slightly variable size in all species tested. A second band of higher molecular weight (asterick in Figure 2A) was only occasionally observed in tomato and tobacco extracts (tobacco not shown in Figures 2A-B) and might represent an aggregate of MFP1. In contrast, a288 antibody only detected a protein of about 80 kD in tomato and

tobacco extracts, suggesting that the DNA-binding domain of MFP1 is more highly conserved than the part of the coiled-coil domain present on fragment E-196.

Together, these data indicate that a protein of similar size, containing a related DNA-binding domain, is conserved among higher plant species and that the highest degree of similarity to tomato (LeMFP1) among the plants investigated can be expected in tobacco.

EXAMPLE 4

Cloning and Characterization Of Several Tobacco MFP1 cDNAs

Corresponding To Two Tobacco MFP1 Proteins

Example 4 describes the cloning and characterization of two, distinct MFP1 proteins from tobacco.

Isolation Of cDNA By Hybridization

The cDNAs encoding tobacco MFP1 were cloned and characterized. An oligo-dT-primed lambda-ZAP cDNA library made from *Nicotiana tabacum* var. SR1 leaf tissue was purchased from Stratagene (La Jolla, CA). The library was screened in a DNA-hybridization screen according to Maniatis with a 1.6 kb partial cDNA clone representing the 3' 2/3 of the tomato homolog of MFP1, LeMFP1 cDNA (p7 2; SEQ ID NO:9) (Meier et al., *Plant Cell* 8:2105-2115 (1996)) or a 1.0 kb 5' partial LeMFP1 cDNA clone (p1-3; SEQ ID NO:10) (Meier et al., *Plant Cell* 8:2105-2115 (1996)). Hybridization conditions were 5 x Denhards (Maniatis), 5 x SSPE (Maniatis), 5% SDS, 20 µg/mL salmon sperm DNA at 55 °C. Washes were performed at high stringency (0.1 x SSC, 0.1 % SDS at 65 °C). Positive plaques were detected by autoradiography and carried through two subsequent rounds of purification, as described above. *In vivo* excision of positive phage was performed according to the manufacturer's protocol (Stratagene, La Jolla, CA).

Sequencing

In the first screen two positive plaque-forming units (pfus) were detected among approximately 600,000 pfus. After *in vivo* excision, sequence analysis of the two excised cDNAs (T6 (SEQ ID NO:11) and T1 (SEQ ID NO:12)) showed that they represented 1103 bp and 912 bp C-terminal MFP1 sequence, respectively (Figure 3). DNA sequencing was carried out using an ABI Model 377 Sequencer (Perkin Elmer-ABI, Foster City, CA). Sequencing reactions utilized fluorescent sequencing techniques with d-rhodamine and Big Dye terminator chemistry (Perkin Elmer-ABI, Foster City, CA) and were performed according to the standard protocols. The sequence identity between the two 3' fragments T1 (SEQ

ID NO:12) and T6 (SEQ ID NO:11) is 91.5%, suggesting the presence of two MFP1 genes in *Nicotiana tabacum*.

In a second round, the tobacco cDNA library was screened with a 1.0 kb 5' fragment of the LeMFP1 cDNA (p1-3; SEQ ID NO:10) (Meier et al., *Plant Cell* 8:2105-2115 (1996)). Two positive pfus were detected among approximately 600,000 pfus. Sequencing of the excised cDNAs (T2 (SEQ ID NO:13) and T3 (SEQ ID NO:14)) showed that they represented partial cDNAs, overlapping with T1 (SEQ ID NO:12) and T6 (SEQ ID NO:11) (Figure 3). Initial sequence analysis of the T2 (SEQ ID NO:13) and T6 (SEQ ID NO:11) cDNAs showed that they shared 445 bp of identical overlapping sequence. It was concluded the T2 (SEQ ID NO:13) and T6 (SEQ ID NO:11) cDNAs represent different portions of the same gene. The overlap of T3 (SEQ ID NO:14) and T1 (SEQ ID NO:12) is only 70 bp, and within this area, there is only a single base pair difference between T6 (SEQ ID NO:11) and T1 (SEQ ID NO:12).

In order to confirm that T3 (SEQ ID NO:14) and T1 (SEQ ID NO:12) were derived from the same gene, PCR primers (SEQ ID NO:25 and SEQ ID NO:26) were designed from the T3 (SEQ ID NO:14) and T1 (SEQ ID NO:12) sequences, that would allow the amplification of a 397 bp fragment, from a *Nicotiana tabacum* lambda ZAP cDNA library, overlapping both cDNAs. PCR reactions were carried out in a Perkin Elmer 9600 thermocycler. (Perkin Elmer, Foster City, CA). The thermocycler was programmed as follows: 2 min 96 °C denaturation cycle, was followed by 30 cycles of 94 °C, 45 sec; 55 °C, 45 sec; 72 °C, 90 sec, and ended with an 8 min 72 °C final extension cycle.

Cloning

Using restriction sites added to the primers, the PCR fragment was subsequently cloned into the XbaI/BamHI sites of pSK+ (Stratagene, La Jolla, CA). The sequence of the fragment PCR1 (SEQ ID NO:15) (Figure 3) was found to be 100% identical with both T1 (SEQ ID NO:12) and T3 (SEQ ID NO:14), confirming that these two cDNA fragments are derived from the same gene.

Figure 3 shows a schematic structure of the partial tobacco MFP1 cDNAs. T3, T1 and PCR1, shown as open boxes, represent overlapping fragments of the same gene (NtMFP1-1). T2 and T6, shown as filled boxes, represent overlapping fragments of a second gene (NtMFP1-2). The fragment used as a probe for the Southern blot (Figure 5) is indicated.

Confirmation Of The Presence Of Two Genes

The divergence between the two tobacco MFP1 cDNAs indicated that they were derived from two different genes. It has been previously shown that a single gene (SEQ ID NO:16) codes for MFP1 (SEQ ID NO:17) in tomato (LeMFP1)

(Meier et al., *Plant Cell* 8:2105-2115 (1996)). Applicants have additionally found that AtMFP1 is a single gene in *Arabidopsis* (data not shown). Based on these findings it was necessary to confirm whether MFP1 is also a single-copy gene in the two diploid progenitors of amphidiploid *Nicotiana tabacum*,

- 5 *N. tomentosiformis* and *N. sylvestris*. In order to confirm this hypothesis the following procedure was applied.

For a Southern blot of *Nicotiana tabacum* genomic DNA, 20 µg aliquots of DNA were digested with various restriction enzymes, run out on 0.8% agarose gel, and were transferred to Immobilon N hydrophobic filters (Millipore, Bedford, MA). Hybridization conditions were essentially as described by Maniatis. The probe (SEQ ID NO:18) was prepared by purification of a 391 bp XhoI/SpeI fragment from the *Nicotiana tabacum* clone T3, as described above. The hybridization temperature was 65 °C. The probe (SEQ ID NO:18), shown in Figure 3, was labelled with ³²P by random prime method according to the manufacturers instructions (BRL, Gaithersburg, MD). Washes were performed at high stringency (0.1 x SSC, 0.1 % SDS at 65 °C).

In the region overlapping the probe, NtMFP1-1 (SEQ ID NO:1) contained a single XbaI site, whereas NtMFP1-2 (SEQ ID NO:3) contains no XbaI site. Neither of the two cDNAs contained an EcoRI site. Figure 5 shows the genomic organization of tobacco MFP1. Tobacco genomic DNA was digested with the indicated restriction enzymes, separated by agarose gel electrophoreses and hybridized in a genomic Southern blot with 391 bp Xho/Spe fragment from the *Nicotiana tabacum* cDNA clone T3 (shown in Figure 3). Abbreviations used are as follows: E, EcoRI; X, XbaI; E/X, EcoRI/XbaI; S, SspI; S/X, SspI/XbaI. The position of DNA size markers is indicated on the right. Two fragments were detected in the lane containing an EcoRI digest (approximately 3.7 kb and 2.7 kb) and three were seen in the lane containing an XbaI digest (approximately 8.0 kb, 7.5 kb and 5.0 kb). In the lane containing the EcoRI/XbaI double digest, the 3.7 kb EcoRI fragment appears to be cleaved by XbaI, leading to two smaller fragments of approximately 1.6 and 0.8 kb. This pattern is consistent with the presence of two genes, one of which contains an XbaI site in the region hybridizing to the probe. In addition, SspI and SspI/XbaI digests were analyzed. Again, one of the two bands detected in the SspI digest is cleaved in the SspI/XbaI double digest. The observed patterns are all consistent with the presence of two genes in the *Nicotiana tabacum* genome, represented by the two isolated cDNAs. These data indicate that, at least in tomato, tobacco and *Arabidopsis*, (data not shown), MFP1 is encoded by a single gene per diploid genome.

In summary, two distinct NtMFP1 cDNAs were isolated from tobacco and named NtMFP1-1 (SEQ ID NO:1) (containing T3 and T1) and NtMFP1-2 (SEQ ID NO:3) (containing T2 and T6). NtMFP1-1 (SEQ ID NO:1) is a full-length cDNA coding for a protein of 722 amino acids (SEQ ID NO:2). NtMFP1-1 (SEQ ID NO:1) and NtMFP1-2 (SEQ ID NO:3) have 77.0% and 78.9% identity to LeMFP1 (SEQ ID NO:16) on DNA level, respectively. The identity between the two tobacco sequences is 91.5%. NtMFP1-1 (SEQ ID NO:2) contains an open reading frame of 721 amino acids. It contains a short 69 bp 5' non-coding region preceding the ATG start codon. NtMFP1-2 (SEQ ID NO:4) contains an open reading frame of 398 amino acids and is not a full-length cDNA.

Table 1 lists the DNASTAR and BLAST comparison of Nt-MFP1-1 and Nt-MFP1-2 with a suit of public databases as well as the literature sequence for tomato (SEQ ID NO:16 and 17) and the MFP1 sequence isolated from Arabidopsis (<http://genomewww.standard.edu/Arabidopsis/>).

15

TABLE I
Comparison of Nt-MPF1-1 and Nt-MPF1-2 Against Known Sequences

Gene/protein	% Identity*		Similarity Identified	BLAST		Citation
	to Tomato	Arabidopsis		Algorithm	E value ¹	
Nt-MFP1-1	<u>AA</u> ²	<u>AA</u> ²	(Y07861) MFP1 protein [<i>Lycopersicon esculentum</i>]	Xnr	0.0c	Meier et al., <i>Plant Cell</i> 8:2105-2115 (1996)
	77%	42%				
Nt-MFP1-2	<u>NA</u> ³	<u>NA</u> ³	(Y07861) MFP1 protein [<i>Lycopersicon esculentum</i>]	Xnr	e-147	Meier et al., <i>Plant Cell</i> 8:2105-2115 (1996)
	79%	49%				

*Comparison made using DNASTAR MEGALIGN, applying default parameters.

¹Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

²AA is the abbreviation for amino acid sequence

³NA is the abbreviation for nucleotide sequence

EXAMPLE 5**Primary And Secondary Structure Analysis Of NtMFP1-1 And NtMFP1-2**

Due to the small number of MFP1-like proteins discovered to date, it was advisable to confirm the identity of the present proteins through an analysis of secondary protein structure. Comparisons of the Nt-MFP1-1 and Nt-MFP1-2 proteins were made with the secondary structure of LeMFP1 isolated from tomato and AtMFP1 isolated from *Arabidopsis*.

The *Arabidopsis* genomic DNA sequence was accessed through the *Arabidopsis thaliana* Database (<http://genomewww.standard.edu/Arabidopsis/>). The deduced protein sequences of the MFP1 proteins were determined and compared using DNASTAR Lasergene software (DNASTAR, Inc., Madison, WI). Figure 4A shows the percent identical amino acids in pairwise comparisons of the four MFP1 proteins.

Based on the amino acid sequence identity NtMFP1-1 (SEQ ID NO:2) and NtMFP1-2 (SEQ ID NO:4) are most closely related. LeMFP1 (SEQ ID NO:17) is more closely related to the two tobacco MFP1s (NtMFP1-1 (SEQ ID NO:2) and NtMFP1-2 (SEQ ID NO:4)) (76% overall sequence identity) than to AtMFP1 (41% overall identity) reflecting the closer relationship of the two solanaeceous species.

Figure 4B shows the hydrophilicity and secondary structure analysis of NtMFP1-1 (SEQ ID NO:2), LeMFP1 (SEQ ID NO:17) and AtMFP1. The secondary structures of the proteins, hydrophilicity, α -helical, and coiled-coil regions were analyzed using DNASTAR PROTEAN software. AH indicates α -helical, CC indicates coiled-coil and HP indicates hydrophilicity plot. The hydrophobic domains are marked with open boxes. Like LeMFP1 (SEQ ID NO:17), NtMFP1 and AtMFP1 contain an extended α -helical, coiled-coil like domain and a shorter N-terminal, non- α -helical region that contains two hydrophobic domains. These structural features are extremely well conserved, despite a relatively low degree of identity on amino acid level in some areas. This is consistent with the more structural conservation of the positioning of polar and non-polar amino acids that is known from other filament-like coiled-coil proteins such as the nuclear lamins (McKeon et al., *Nature* 319:463-468 (1986)). The distance between the first and second hydrophobic domains is very similar in all three proteins (29 AA for tomato, 31 AA for tobacco, and 33 AA for *Arabidopsis* MFP1), indicating a functional relevance of the spacing between the two hydrophobic domains. The length of the N-terminal domain preceding the first hydrophobic domain varies between 56 AA for tomato, 61 AA for tobacco, and 72 AA for *Arabidopsis* MFP1. The common feature of this domain in all three

proteins is a relatively high content of serine and threonine residues (27% to 28%).

EXAMPLE 6

Composition Of cDNA Libraries And Identification

5 Of cDNA Clones From Other Plant Species Encoding Homologs Of MFP1

cDNA libraries representing mRNAs from soybean or corn tissues were prepared. The characteristics of the libraries are described below in Table 2.

Table 2

cDNA Libraries from Plants	Library	Tissue
Soybean (<i>Glycine max</i>)	src3c.pk004.m1	8 day old root tissue inoculated with eggs of nematode
Corn (<i>Zea mays</i>)	p0118.chsab48r	stem tissue, night harvested
Rice (<i>Oryza sativa</i> L., Nipponbare)	rcal.n.pk022.a11	callus normalized

10

Soybean MFP1:

A soybean MFP1 cDNA was identified based on primary and secondary structure analysis. This sequence, from clone src3c.pk004.m1, came from a library prepared from 8 day old root tissue inoculated with eggs of cyst nematode for four days. This sequence contains 1164 base pairs of DNA (SEQ ID NO:19) encoding 388 amino acids (SEQ ID NO:20).

Comparison of this partial soybean MFP1 sequence (SEQ ID NO:20) with the sequences from tomato (LeMFP1; SEQ ID NO:17), tobacco (NtMFP1-1; SEQ ID NO:2), and *Arabidopsis* (AtMFP1) shows it to be 46.1, 45.9, and 40.7% identical to these sequences, respectively. In addition, secondary structure analysis of the partial soybean MFP1 (GmMFP1; SEQ ID NO:20) coded by SEQ ID NO:19 shows that it contains an extended α -helical, coiled-coil like domain as do the other MFP1 protein sequences. Results of a Southern blot experiment (not shown) suggest that the soybean MFP1 is encoded by a single copy gene.

25 Corn MFP1:

A corn EST sequence was identified as an MFP1 homolog from clone p0118.chsab48r. Secondary structure analysis of the corn MFP1 protein (SEQ ID NO:22) coded by SEQ ID NO:21 shows that it contains an extended α -helical, coiled-coil like domain as well as one of the hydrophobic domains at the N-terminus. Both features are indicative of MFP1 proteins (see Figure 4B).

30

Rice MFP1:

A rice EST, from clone rca1n.pk022.a11, was isolated which codes for an MFP1 protein. The identity of the rice EST was based on its high degree of identity to the corn MFP1 sequence (68%). This clone covers the C-terminal
5 region that is most highly conserved between all MFP1 proteins identified. The rice MFP1 sequence (SEQ ID NO:23) codes for SEQ ID NO:24.

Table 3 lists the DNASTAR and BLAST comparison of the MFP1 sequences isolated from corn, soybean and rice with a suit of public databases as well as the literature sequence for tomato (SEQ ID NO:16 and 17) and the
10 sequence isolated from Arabidopsis
(<http://genomewww.standard.edu/Arabidopsis/>).

TABLE 3
Comparison of Corn, Rice and Soybean MFP1-1 Sequences Against Known Sequences

Gene/protein	% Identity* to		% Identity* to		Similarity Identified	BLAST Algorithm	E value ¹	Citation
	AA ²	NA ³	AA ²	NA ³				
Corn MFP1	40%	40%	34%	48%	(Y07861) MFP1 protein [<i>Lycopersicon</i> <i>esculentum</i>]	Xnr	2e-94	Meier et al., <i>Plant Cell</i> 8:2105-2115 (1996)
Soybean MFP1	46%	51%	41%	48%	(Y07861) MFP1 protein [<i>Lycopersicon</i> <i>esculentum</i>]	Xnr	5e-76	Meier et al., <i>Plant Cell</i> 8:2105-2115 (1996)
Rice MFP1	39%	43%	31%	39%	(Y07861) MFP1 protein [<i>Lycopersicon</i> <i>esculentum</i>]	Xnr	6e-12	Meier et al., <i>Plant Cell</i> 8:2105-2115 (1996)

*Comparison made using DNASTAR MEGALIGN, applying default values

¹Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

²AA is the abbreviation for amino acid sequence

³NA is the abbreviation for nucleotide sequence

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding a plant MFP1 protein selected from the group consisting of:
 - 5 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24;
 - 10 (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24;
 - 15 (c) an isolated nucleic acid molecule that hybridizes with a nucleic acid sequence of (a) or (b) under the following hybridization conditions: 5 x Denhards, 5 x SSPE, 5% SDS, 20 µg/mL salmon sperm DNA at 55 °C;
 - 20 (d) an isolated nucleic acid molecule that hybridizes with a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21 and SEQ ID NO:23 under the following hybridization conditions: 5 x Denhards, 5 x SSPE, 5% SDS, 20 µg/mL salmon sperm DNA at 55 °C; and
 - 25 (e) an isolated nucleic acid fragment that is complementary to (a), (b), (c) or (d).
2. The isolated nucleic acid fragment of Claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21 and SEQ ID NO:23.
- 30 3. A polypeptide encoded by the isolated nucleic acid fragment of Claim 1.
4. The polypeptide of Claim 3 selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24.
- 35 5. A nucleic acid fragment, isolated from corn, encoding an MFP1 polypeptide, the polypeptide having at least 40% identity to SEQ ID NO:17, over a length of about 672 amino acids as compared by the Jotun-Hein algorithm.

6. An MFP1 polypeptide encoded by the nucleic acid fragment of Claim 5.
7. A nucleic acid fragment, isolated from soybean, encoding an MFP1 polypeptide, the polypeptide having at least 46% identity to SEQ ID NO:17 over a length of 388 amino acids as compared by the Jotun-Hein algorithm.
8. An MFP1 polypeptide, encoded by the nucleic acid fragment of Claim 7.
9. A nucleic acid fragment, isolated, from rice, encoding an MFP1 polypeptide, the polypeptide having at least 39% identity to SEQ ID NO:17 over a length of 107 amino acids as compared by the Jotun-Hein algorithm.
10. An MFP1 polypeptide, encoded by the nucleic acid fragment of Claim 9.
11. An isolated nucleic acid fragment encoding a plant MFP1 polypeptide, the peptide having at least 77% identity to SEQ ID NO:17.
12. An MFP1 polypeptide encoded by the nucleic acid fragment of Claim 11.
13. A chimeric gene comprising the isolated nucleic acid fragment of any of Claims 1, 5, 7, 9, and 11 operably linked to suitable regulatory sequences.
14. A transformed host cell comprising a host cell and the chimeric gene of Claim 13.
15. The transformed host cell of Claim 14 wherein the host cell is a plant cell.
16. The transformed host cell of Claim 14 wherein the host cell is *E. coli*.
17. A method of altering the level of expression of a plant MFP1 protein in a host cell comprising:
 - (a) transforming a host cell with the chimeric gene of Claim 13 and;
 - (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene resulting in production of altered levels of a plant MFP1 protein in the transformed host cell relative to expression levels of an untransformed host cell.
18. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a plant MFP1 protein comprising:
 - (a) probing a cDNA or genomic library with the nucleic acid fragment of Claim 1;
 - (b) identifying a DNA clone that hybridizes with the nucleic acid fragment of Claim 1; and

- (c) sequencing the cDNA or genomic fragment that comprises the clone identified in step (b),

wherein the sequenced cDNA or genomic fragment encodes a plant MFP1 protein

19. A method of obtaining a nucleic acid fragment encoding all or a
5 substantial portion of the amino acid sequence encoding a plant MFP1 protein comprising:

- 10 (a) synthesizing at least one oligonucleotide primer corresponding to a portion of the sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21 and SEQ ID NO:23;
- (b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a);

wherein the amplified cDNA insert encodes a plant MFP1 protein.

- 15 20. The product of the method of Claims 18 or 19.

1/4

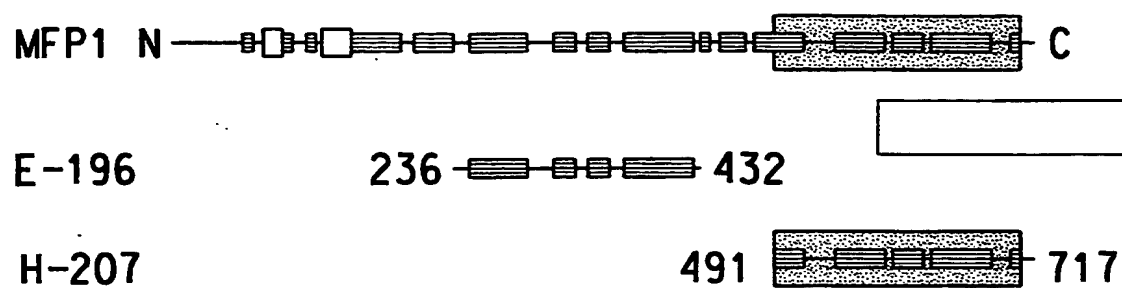


FIG. 1

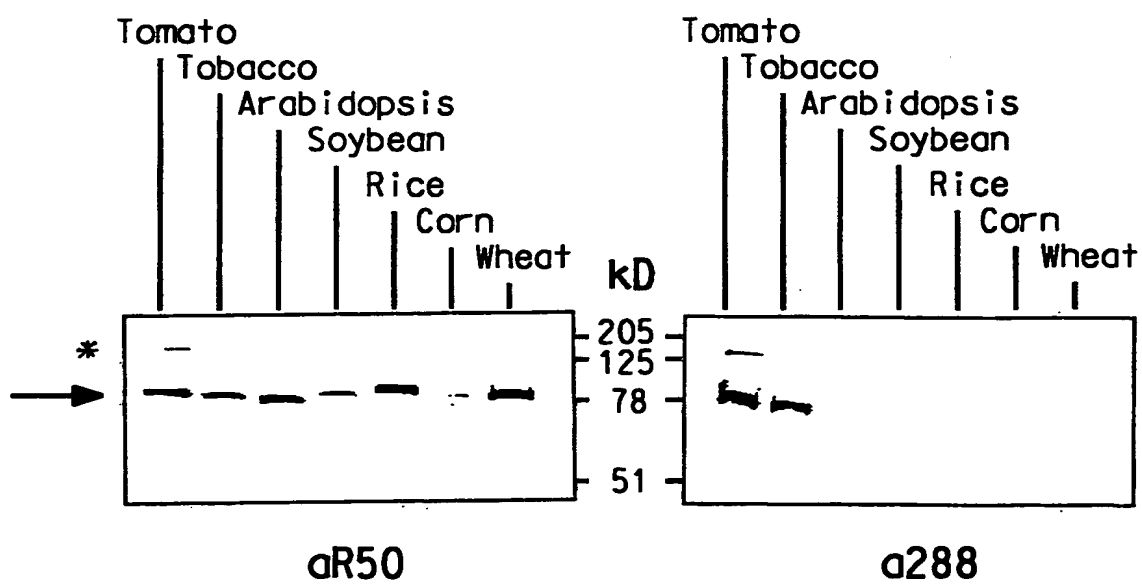


FIG 2A

FIG. 2B

2/4

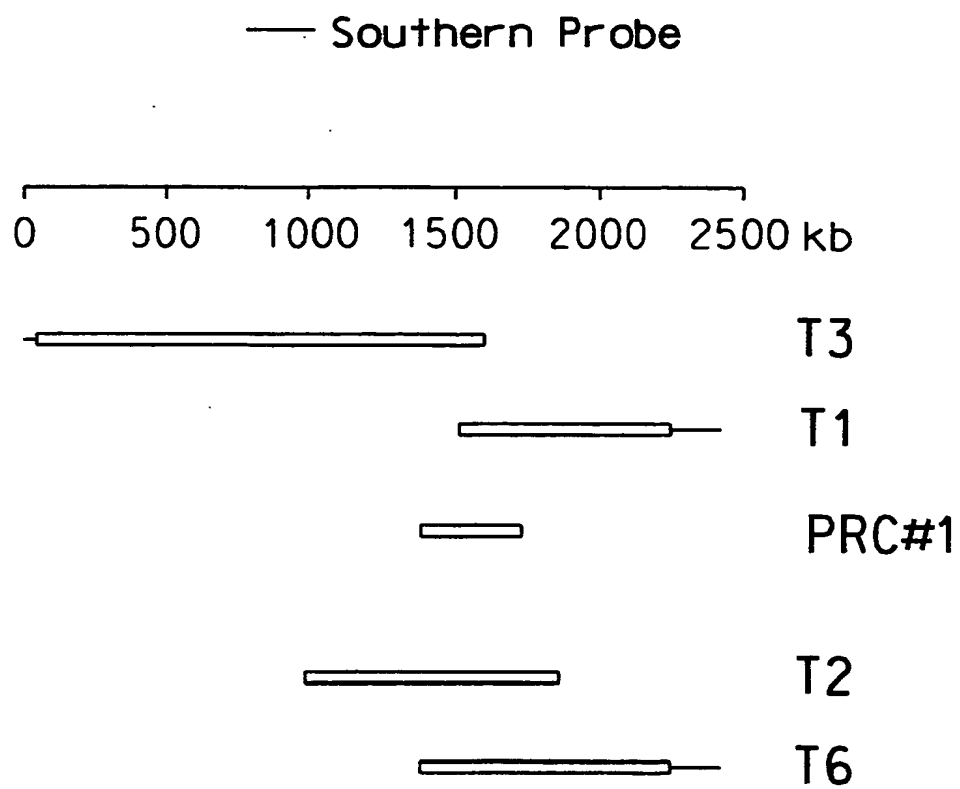


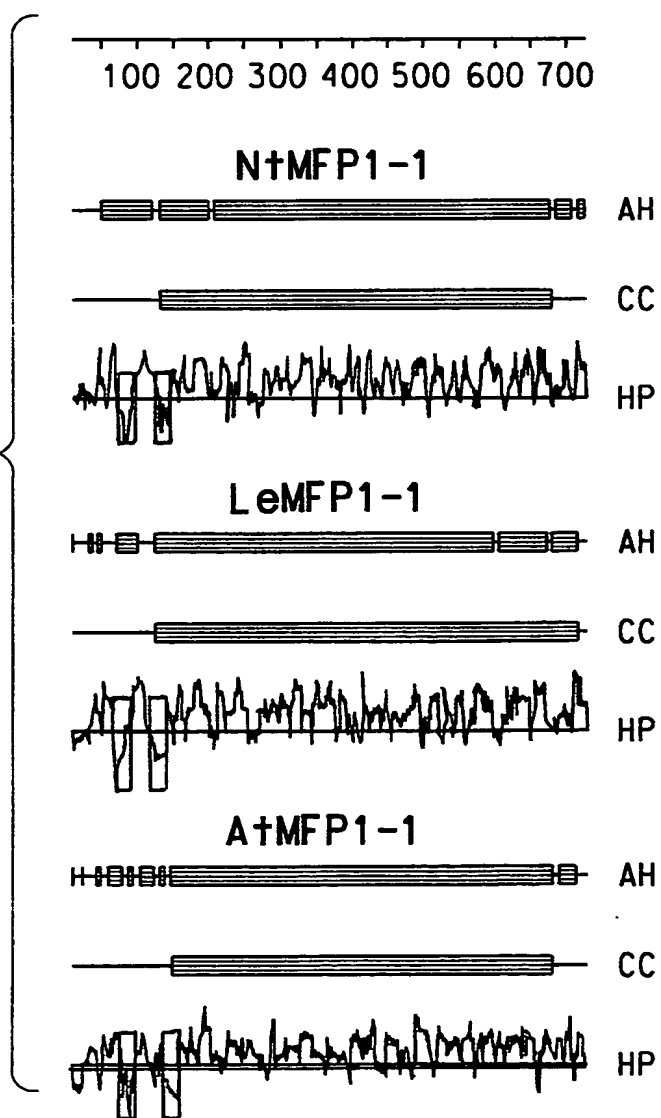
FIG. 3

3/4

	N+MFP1-1	N+MFP1-2	LeMFP1	A+MFP1
N+MFP1-1		91.5	77.0	41.9
N+MFP1-2			78.9	42.2
LeMFP1				41.5

FIG. 4A

FIG. 4B



4/4

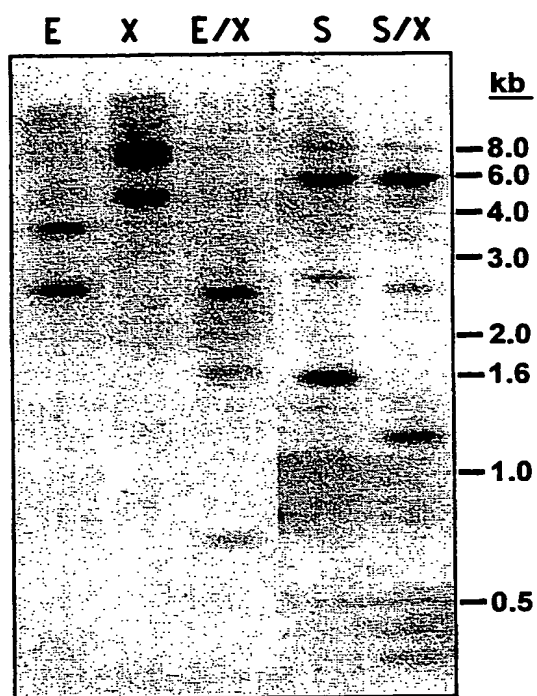


FIG. 5

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Homologs of MAR-binding Filament-like protein 1 (MFPl)

<130> BC1003 PCT

<140>

<141>

<150> 60/128,900

<151> 1999-04-12

<160> 26

<170> Microsoft Office 97

<210> 1

<211> 2168

<212> DNA

<213> Nicotiana tabacum

<400> 1

atggggagtt	cttgttttcc	ccaatctcca	ctctctcatt	ctctcttttc	ttcttcatca	60
atatcttctt	cccaatttac	acccttgctt	ttttcccca	gaaatgcgca	aaaatgtaaa	120
aagaaaatgc	cagctatggc	atgtatacac	tcgagaaatc	aaaaggaaag	cgaattctgc	180
agcagaagaa	cgattctttt	cgtgggtttc	tctgttcttc	cacttctcag	cttgagggca	240
aatgcttttg	aaggcttgtc	agtagattct	caagtaaaag	cacagccgca	gaaagaggag	300
acagagcaaa	caatccaagg	aaatgcagag	aatcccttct	tttctctact	taatggactt	360
ggagtttttg	gttcaggcgt	gcttggttct	ctttatgcct	tggctcgaaa	cgagaaggcc	420
gtttctgatg	caaccattga	atctatgaaa	aataagctga	aggagaaaga	agccacattc	480
gtttcatgga	gaagaaattc	cagtctgagc	tgctgaacga	aagggatata	cgaaataatc	540
aacttaagag	ggcaggcgaa	gaacggcaag	ctctgggtta	ccaattgaat	tcagcaaaga	600
gtacagtaac	taaccttggt	caggagctgc	aaaaagaaaa	acgaattgct	gaagagctca	660
tagttcagat	cgagggcctt	caaaataacc	tcatgcagat	gaaggaggat	aagaaaaaat	720
tgcaggagga	gcttaaagag	aagcttgatt	tgatacaagt	tctgcaagaa	aagataactt	780
tacttactac	agagatcaaa	gataaagagg	catctcttca	gagtacaacc	tctaaactag	840
ctgaaaaaga	atcagaggta	gataaattga	gctcaatgta	tcaggaatcc	caggatcagc	900
tgatgaattt	gacttcagaa	atcaaagaac	ttaaagtcga	agtccagaaa	agagagagag	960
aactagagtt	gaaacgtgaa	tcagaagaca	accttaatgt	gcgattaaat	tctttgctcg	1020
ttgagagaga	tgaatctaaa	aaagagcttg	atgctattca	aaaggaatac	agcgagttca	1080
agtcattttc	agagaagaaa	gtggcttctg	atgccaaagt	gttgggggaa	caagaaaaga	1140
gactacacca	gctcgaggaa	caacttggca	ctgcctcaga	tgaagtacgc	aaaaataatg	1200
tgctaatcgc	tgatctgact	caagaaaaag	aaaacttaag	gagaatgctg	gacgctgagc	1260
tggaaaacat	aagcaagttg	aagctagagg	tccaggttac	tcaggaaact	cttgagaaat	1320
ctagaagtga	tgcttctgat	atagcacaac	aactacagca	gtcgaggcat	ctttgctcta	1380
agcttgaagc	tgaggtttct	aaacttcaga	tggaattgga	ggaaacaaga	acatcattac	1440
ggaggaacat	tgatgagaca	aaacgtggtg	cagagctctt	agctgcggag	ctgaccacta	1500
ctagggagct	tctaaagaaa	acaaatgaag	aaatgcacac	tatgtctcat	gaactagcgg	1560
ctgttactga	aaattgtgat	aacttacaga	cggagctagt	tgatgtctac	aagaaagcag	1620
aacgtgctgc	tgatgaactg	aaacaagaaa	agaatattgt	cgtgacactg	gagaaagagc	1680
taacattttt	ggaggctcaa	attacaagag	agaaagagtc	acggaagaat	ctggaagaag	1740
agctggaaaag	ggctacggaa	tcacttgatg	agatgaaccg	aaatgctttt	gcacttgcaa	1800
aggagcttga	gcttgctaat	tctcatattt	ctagcctcga	ggatgagaga	gaagtgtctc	1860
aaaagtctgt	ttctgagcag	aaacaaattt	ctcaagaatc	ccgagaaaac	cttgaagatg	1920
cccatagcct	ggtaatgaaa	cttggcaagg	aacgcgagag	tctggagaag	agagcaaaga	1980
aattggaaga	tgaaatggca	tcagcaaaaag	gtgagatttt	gcggtgctgc	acccaagtaa	2040
attcggtaaa	agctcctgtt	aacaatgagg	aaaaagttga	agctggggaa	aaggcagctg	2100
taacagtga	gagaaccagg	aggaggaaga	ctgctactca	gcctgcttct	cagcaagaaa	2160
gctcatag						2168

<210> 2

<211> 721

<212> PRT

<213> Nicotiana tabacum

<400> 2

Met Gly Ser Ser Cys Phe Pro Gln Ser Pro Leu Ser His Ser Leu Phe
1 5 10 15

Ser Ser Ser Ser Ile Ser Ser Ser Gln Phe Thr Pro Leu Leu Phe Ser
20 25 30

Pro Arg Asn Ala Gln Lys Cys Lys Lys Lys Met Pro Ala Met Ala Cys
35 40 45

Ile His Ser Glu Asn Gln Lys Glu Ser Glu Phe Cys Ser Arg Arg Thr
50 55 60

Ile Leu Phe Val Gly Phe Ser Val Leu Pro Leu Leu Ser Leu Arg Ala
65 70 75 80

Asn Ala Phe Glu Gly Leu Ser Val Asp Ser Gln Val Lys Ala Gln Pro
85 90 95

Gln Lys Glu Glu Thr Glu Gln Thr Ile Gln Gly Asn Ala Glu Asn Pro
100 105 110

Phe Phe Ser Leu Leu Asn Gly Leu Gly Val Phe Gly Ser Gly Val Leu
115 120 125

Gly Ser Leu Tyr Ala Leu Ala Arg Asn Glu Lys Ala Val Ser Asp Ala
130 135 140

Thr Ile Glu Ser Met Lys Asn Lys Leu Lys Glu Lys Glu Ala Thr Phe
145 150 155 160

Val Ser Met Glu Lys Lys Phe Gln Ser Glu Leu Leu Asn Glu Arg Asp
165 170 175

Ile Arg Asn Asn Gln Leu Lys Arg Ala Gly Glu Glu Arg Gln Ala Leu
180 185 190

Val Asn Gln Leu Asn Ser Ala Lys Ser Thr Val Thr Asn Leu Gly Gln
195 200 205

Glu Leu Gln Lys Glu Lys Arg Ile Ala Glu Glu Leu Ile Val Gln Ile
210 215 220

Glu Gly Leu Gln Asn Asn Leu Met Gln Met Lys Glu Asp Lys Lys Lys
225 230 235 240

Leu Gln Glu Glu Leu Lys Glu Lys Leu Asp Leu Ile Gln Val Leu Gln
245 250 255

Glu Lys Ile Thr Leu Leu Thr Thr Glu Ile Lys Asp Lys Glu Ala Ser
260 265 270

Leu Gln Ser Thr Thr Ser Lys Leu Ala Glu Lys Glu Ser Glu Val Asp
275 280 285

Lys Leu Ser Ser Met Tyr Gln Glu Ser Gln Asp Gln Leu Met Asn Leu
290 295 300

Thr Ser Glu Ile Lys Glu Leu Lys Val Glu Val Gln Lys Arg Glu Arg
305 310 315 320

Glu	Leu	Glu	Leu	Lys	Arg	Glu	Ser	Glu	Asp	Asn	Leu	Asn	Val	Arg	Leu	
				325					330					335		
Asn	Ser	Leu	Leu	Val	Glu	Arg	Asp	Glu	Ser	Lys	Lys	Glu	Leu	Asp	Ala	
			340					345						350		
Ile	Gln	Lys	Glu	Tyr	Ser	Glu	Phe	Lys	Ser	Ile	Ser	Glu	Lys	Lys	Val	
		355					360					365				
Ala	Ser	Asp	Ala	Lys	Leu	Leu	Gly	Glu	Gln	Glu	Lys	Arg	Leu	His	Gln	
	370					375					380					
Leu	Glu	Glu	Gln	Leu	Gly	Thr	Ala	Ser	Asp	Glu	Val	Arg	Lys	Asn	Asn	
385					390					395					400	
Val	Leu	Ile	Ala	Asp	Leu	Thr	Gln	Glu	Lys	Glu	Asn	Leu	Arg	Arg	Met	
				405					410					415		
Leu	Asp	Ala	Glu	Leu	Glu	Asn	Ile	Ser	Lys	Leu	Lys	Leu	Glu	Val	Gln	
			420					425					430			
Val	Thr	Gln	Glu	Thr	Leu	Glu	Lys	Ser	Arg	Ser	Asp	Ala	Ser	Asp	Ile	
		435					440					445				
Ala	Gln	Gln	Leu	Gln	Gln	Ser	Arg	His	Leu	Cys	Ser	Lys	Leu	Glu	Ala	
	450					455					460					
Glu	Val	Ser	Lys	Leu	Gln	Met	Glu	Leu	Glu	Glu	Thr	Arg	Thr	Ser	Leu	
465					470					475					480	
Arg	Arg	Asn	Ile	Asp	Glu	Thr	Lys	Arg	Gly	Ala	Glu	Leu	Leu	Ala	Ala	
				485					490					495		
Glu	Leu	Thr	Thr	Thr	Arg	Glu	Leu	Leu	Lys	Lys	Thr	Asn	Glu	Glu	Met	
			500					505					510			
His	Thr	Met	Ser	His	Glu	Leu	Ala	Ala	Val	Thr	Glu	Asn	Cys	Asp	Asn	
		515					520					525				
Leu	Gln	Thr	Glu	Leu	Val	Asp	Val	Tyr	Lys	Lys	Ala	Glu	Arg	Ala	Ala	
	530					535					540					
Asp	Glu	Leu	Lys	Gln	Glu	Lys	Asn	Ile	Val	Val	Thr	Leu	Glu	Lys	Glu	
545					550					555					560	
Leu	Thr	Phe	Leu	Glu	Ala	Gln	Ile	Thr	Arg	Glu	Lys	Glu	Ser	Arg	Lys	
				565					570					575		
Asn	Leu	Glu	Glu	Glu	Leu	Glu	Arg	Ala	Thr	Glu	Ser	Leu	Asp	Glu	Met	
			580					585					590			
Asn	Arg	Asn	Ala	Phe	Ala	Leu	Ala	Lys	Glu	Leu	Glu	Leu	Ala	Asn	Ser	
		595					600					605				
His	Ile	Ser	Ser	Leu	Glu	Asp	Glu	Arg	Glu	Val	Leu	Gln	Lys	Ser	Val	
	610					615					620					
Ser	Glu	Gln	Lys	Gln	Ile	Ser	Gln	Glu	Ser	Arg	Glu	Asn	Leu	Glu	Asp	
625					630					635					640	
Ala	His	Ser	Leu	Val	Met	Lys	Leu	Gly	Lys	Glu	Arg	Glu	Ser	Leu	Glu	
				645					650					655		

Lys Arg Ala Lys Lys Leu Glu Asp Glu Met Ala Ser Ala Lys Gly Glu
660 665 670

Leu Arg Leu Arg Thr Gln Val Asn Ser Val Lys Ala Pro Val Asn Asn
675 680 685

Glu Glu Lys Val Glu Ala Gly Glu Lys Ala Ala Val Thr Val Lys Arg
690 695 700

Thr Arg Arg Arg Lys Thr Ala Thr Gln Pro Ala Ser Gln Gln Glu Ser
705 710 715 720

Ser

<210> 3
<211> 1199
<212> DNA
<213> Nicotiana tabacum

<400> 3
cgagatgtga atcagaagac aacctgaatg tgcaattaaa ttctttgctc gttgagagag 60
atgaatctaa aaaagagctt gatgctattc aaaaggaata cagcgagttc agtccattt 120
cagagaagag agtggcttca gatgccaagc tggtggggga acaagaaaag agactacacc 180
agctcgagga acaacttggg actgccgtaa gtgaagtaag aaaaaataaa gtgctaattg 240
ctaatttgac tcaagcaaaa gaaaacctaa ggagaatgct ggacgctgag ctggaaaatg 300
taagcaagtt gaagctagag gtccagggtta ctcaggaaac tcttgagaaa tcaagaagtg 360
aagcttctga tatagtagaa caactacagc agtcgaggca cttgtgctct aagcttgaag 420
ctgaggtttc taagcttcag atggaattgg aggaaacaag gacattgtta cagaagaaca 480
ttgatgagac aaaacgtggg gcagagttct tagctgcgga gctgaccact actagggagc 540
ttctaaagaa aacaaatgaa gaaatgcaca ccatatccaa tgaactagct gctgttactg 600
aaaatcgtga taacttacag acggagctag ttgatgtcta caagaaagca gaacgtgctg 660
ttaatgaact gaaacaagaa aagaatattg tcgtgacatt ggagaaagag ctaacatttt 720
tgagggtctca aattacaaga gagaaagagt cacggaagaa tctggaagaa gagttggaaa 780
gggctacaga atcacttgat gagatgaaca gaaatgcttt tgcacttgca aaggagctgg 840
agctcgctaa ttctcgtatt tctagcctca aagacgagag agaagtgtc caaaagtctg 900
tttctgagca gaagcaaatt tctcaagaag cccgagaaaa ccttgaagat gcccatagcc 960
tggtgatgaa acttggcaag gaacgcgaga gtctggagaa gagagcaaag aaattggaag 1020
atgaaatggc atcagcaaaa ggtgagattt tgcggttgcg gacacaagta aattcggtaa 1080
aagctcctgt taacaaagag gaaaaagttg aagctgggga aaaggcaaca gtaacagtga 1140
agagaacaac caggaggagg aagactgcta ctctgcttc tcaacaagaa ggctcataa 1199

<210> 4
<211> 398
<212> PRT
<213> Nicotiana tabacum

<400> 4
Arg Cys Glu Ser Glu Asp Asn Leu Asn Val Gln Leu Asn Ser Leu Leu
1 5 10 15

Val Glu Arg Asp Glu Ser Lys Lys Glu Leu Asp Ala Ile Gln Lys Glu
20 25 30

Tyr Ser Glu Phe Lys Ser Ile Ser Glu Lys Arg Val Ala Ser Asp Ala
35 40 45

Lys Leu Leu Gly Glu Gln Glu Lys Arg Leu His Gln Leu Glu Glu Gln
50 55 60

Leu Gly Thr Ala Val Ser Glu Val Arg Lys Asn Lys Val Leu Ile Ala
65 70 75 80

```
<210> 5
<211> 588
<212> DNA
<213> Lycopersicon esculentum
```

<400> 5
 agagcttaaa gagaagcttg atttgattca agttcttgaa gaaaagatta ctttgcttac 60
 tacagagatc aaagataaag aggtgagtct tcggagtaac acctctaaac tagctgaaaa 120
 agaatcggag gtaaatagtt tgagcgatat gtatcaacaa tcccaggatc agctgatgaa 180
 tttgacttca gagatcaaag aacttaaaga tgaaatccag aaaagagaga gagaactgga 240
 gttgaaatgt gtatcagaag acaacctgaa tgtgcaatta aattctttgc tcctcgagag 300
 agatgaatct aaaaaagagc ttcatgctat tcaaaaaggaa tacagtgagt tcaagtccaa 360
 ttctgatgag aaggtggctt cagatgcgaa gctgttgggg gaacaagaga agagactaca 420
 ccagcttgag gaacaacttg gcactgcctt aagtgaagca agtaaaaatg aagtgctaata 480
 tgctgatctg actcgagaaa aagaaaacct taggagaatg gtggatgctg agctggacaa 540
 tgtaaacaa g ttaaagcaag agattgaagt cactcaggaa agtcttga 588

<210> 6
 <211> 195
 <212> PRT
 <213> *Lycopersicon esculentum*

<400> 6
 Glu Leu Lys Glu Lys Leu Asp Leu Ile Gln Val Leu Glu Glu Lys Ile
 1 5 10 15
 Thr Leu Leu Thr Thr Glu Ile Lys Asp Lys Glu Val Ser Leu Arg Ser
 20 25 30
 Asn Thr Ser Lys Leu Ala Glu Lys Glu Ser Glu Val Asn Ser Leu Ser
 35 40 45
 Asp Met Tyr Gln Gln Ser Gln Asp Gln Leu Met Asn Leu Thr Ser Glu
 50 55 60
 Ile Lys Glu Leu Lys Asp Glu Ile Gln Lys Arg Glu Arg Glu Leu Glu
 65 70 75 80
 Leu Lys Cys Val Ser Glu Asp Asn Leu Asn Val Gln Leu Asn Ser Leu
 85 90 95
 Leu Leu Glu Arg Asp Glu Ser Lys Lys Glu Leu His Ala Ile Gln Lys
 100 105 110
 Glu Tyr Ser Glu Phe Lys Ser Asn Ser Asp Glu Lys Val Ala Ser Asp
 115 120 125
 Ala Lys Leu Leu Gly Glu Gln Glu Lys Arg Leu His Gln Leu Glu Glu
 130 135 140
 Gln Leu Gly Thr Ala Leu Ser Glu Ala Ser Lys Asn Glu Val Leu Ile
 145 150 155 160
 Ala Asp Leu Thr Arg Glu Lys Glu Asn Leu Arg Arg Met Val Asp Ala
 165 170 175
 Glu Leu Asp Asn Val Asn Lys Leu Lys Gln Glu Ile Glu Val Thr Gln
 180 185 190
 Glu Ser Leu
 195

<210> 7
 <211> 662
 <212> DNA
 <213> *Lycopersicon esculentum*

<400> 7
gaccactact aaggagcttc taaagaaaac aaatgaagaa atgcacacta tgtcagatga 60
accgtgatag cttacagaca gagctagttg atgtctataa gaaagcagaa catactgcta 120
atgaactgaa acaagaaaag agcattgttg caacactaga agaagagtta aaatttctgg 180
agtctcaaatac tacacgagag aaagagttac ggaagagtct ggaagacgag ttagaaaagg 240
ctacagaatc tcttgatgag attaaccgaa atgtgttggc acttgacagag gagctggagc 300
ttgctacttc tcgtaattct agcctcgaag acgagagaga agtgctccga cagtctgttt 360
ctgagcagaa gcaaatttca caagaagccc aagaaaatct ggaagacgcc catagcctgg 420
tgatgaaact tggcaaggaa cgcgaaagtc ttgagaagag agcaaagaaa ttggaagatg 480
aaatggcagc agcaaaaagg gagattttgc ggctacggag ccaaataaac tcagtaaaag 540
ctccagtgga ggatgaggaa aaagttgttg ctggggaaaa ggaaaagggtg aaggcaacag 600
taacagcaaa gaaaactacc aggagaagga agagtgtctac tgttaagcaa gaggaaccct 660
ag

<210> 8
<211> 226
<212> PRT
<213> Lycopersicon esculentum

<400> 8
Thr Thr Thr Lys Glu Leu Leu Lys Lys Thr Asn Glu Glu Met His Thr
1 5 10 15
Met Ser Asp Glu Leu Val Ala Val Ser Glu Asn Arg Asp Ser Leu Gln
20 25 30
Thr Glu Leu Val Asp Val Tyr Lys Lys Ala Glu His Thr Ala Asn Glu
35 40 45
Leu Lys Gln Glu Lys Ser Ile Val Ala Thr Leu Glu Glu Glu Leu Lys
50 55 60
Phe Leu Glu Ser Gln Ile Thr Arg Glu Lys Glu Leu Arg Lys Ser Leu
65 70 75 80
Glu Asp Glu Leu Glu Lys Ala Thr Glu Ser Leu Asp Glu Ile Asn Arg
85 90 95
Asn Val Leu Ala Leu Ala Glu Glu Leu Glu Leu Ala Thr Ser Arg Asn
100 105 110
Ser Ser Leu Glu Asp Glu Arg Glu Val Leu Arg Gln Ser Val Ser Glu
115 120 125
Gln Lys Gln Ile Ser Gln Glu Ala Gln Glu Asn Leu Glu Asp Ala His
130 135 140
Ser Leu Val Met Lys Leu Gly Lys Glu Arg Glu Ser Leu Glu Lys Arg
145 150 155 160
Ala Lys Lys Leu Glu Asp Glu Met Ala Ala Ala Lys Gly Glu Ile Leu
165 170 175
Arg Leu Arg Ser Gln Ile Asn Ser Val Lys Ala Pro Val Glu Asp Glu
180 185 190
Glu Lys Val Val Ala Gly Glu Lys Glu Lys Val Lys Ala Thr Val Thr
195 200 205
Ala Lys Lys Thr Thr Arg Arg Arg Lys Ser Ala Thr Val Lys Gln Glu
210 215 220
Glu Pro
225

<210> 9
 <211> 1694
 <212> DNA
 <213> *Lycopersicon esculentum*

<400> 9
 gaagagctta aagagaagct tgatttgatt caagttcttg aagaaaagat tactttgctt 60
 actacagaga tcaaagataa agagggtgagt cttcggagta acacctctaa actagctgaa 120
 aaagaatcgg aggtaaatag tttgagcgat atgtatcaac aatcccagga tcagctgatg 180
 aatttgactt cagagatcaa agaacttaaa gatgaaatcc agaaaagaga gagagaactg 240
 gagttgaaat gtgtatcaga agacaacctg aatgtgcaat taaattcttt gtcctcagag 300
 agagatgaat ctaaaaaaga gcttcatgct attcaaaagg aatacagtga gttcaagtcc 360
 aattctgatg agaaggtggc ttcagatgag aagctgttg gggaacaaga gaagagacta 420
 caccagcttg aggaacaact tggcactgcc ttaagtgaag caagtaaaaa tgaagtgcta 480
 attgtgatc tgactcgaga aaaagaaaac cttaggagaa tgggtgatgc tgagctggac 540
 aatgtaaaaca agttaagca agagattgaa gtcactcagg aaagtcttga gaattcaaga 600
 agtgaagttt ctgatataac agtacaacta gagcagttga gggatctttg ctccaaactt 660
 gaagctgagg tttctaaact tcagatggaa ttggaggaaa caagggcatc attacagagg 720
 aacattgatg aaacaaaaca cagttcagag ctcttagctg ctgagttgac cactactaag 780
 gagcttctaa agaaaacaaa tgaagaaatg cacactatgt cagatgaact agtagctgtt 840
 tctgaaaatc gtgatagctt acagacagag ctagtgtatg tctataagaa agcagaactg 900
 actgctaagt aactgaaaca agaaaagagc attgttgcaa cactagaaga agagttaaaa 960
 tttctggagt ctcaaattac acgagagaaa gagttacgga agagtctgga agacgagtta 1020
 gaaaaggcta cagaatctct tgatgagatt aaccgaaatg tgttggcact tgagaggag 1080
 ctggagcttg ctacttctcg taattctagc ctcgaagacg agagagaagt gtcggacag 1140
 tctgtttctg agcagaagca aatttcacaa gaagcccaag aaaatctgga agacgccccat 1200
 agcctggtga tgaaacttgg caaggaacgc gaaagtcttg agaagagagc aaagaaattg 1260
 gaagatgaaa tggcagcagc aaaaggtgag attttgcggc tacggagcca aataaactca 1320
 gtaaaagctc cagtggagga tgaggaaaaa gttgttgctg gggaaaagga aaaggtgaag 1380
 gcaacagtaa cagcaaagaa aactaccagg agaaggaaga gtgctactgt taagcaagag 1440
 gaaccctagt tggctgtttc tgaatgacat aatcttcttc tttttttgtc ctgactcatt 1500
 tgtttgcaat atttatagag aggccagaat taggacattg ccatttgaac aagctgtgta 1560
 ttgtctcttt gagtgtacat ttcccggcga gaagttgcag aaacaaatga ctgatctctt 1620
 gatattcagt caatgttgca gcttactgaa tgaaattatt tgtattgtaa aaaaaaaaaa 1680
 aaaaaaaaaa aaaa 1694

<210> 10
 <211> 1009
 <212> DNA
 <213> *Lycopersicon esculentum*

<400> 10
 taataatggc aacttcttgt tttctccat tttctgcttc atcttcttca ttatgttctt 60
 cccaattttac acctttgctt tcttgcccaa gaaataccca aatatgtaga aagaagagac 120
 cggttatggc gagtatgcac tcggaaaatc aaaaggaaaag taatgtctgc aacagaagat 180
 cgattctatt tgtgggattc tcagttcttc cacttctcaa tttgagggca agagctctcg 240
 aaggcttgtc aacagattct caagcacagc cgcagaaaga ggaaaccgag caaacaatcc 300
 aaggaagtgc agggaaatccc ttcgtttctc tacttaatgg acttgggtgtt gttgggtcag 360
 gcgctgttgg ttctctttat gccttggctc gaaatgagaa ggcagtttca gatgcaacca 420
 ttgaatctat gaaaaataag ctgaaggaca aggaagatgc atttgtttca atgaagaagc 480
 aatttgagtc cgaattgctg agcgaaaggg aagatcgaaa taagctaatt aggcgagaag 540
 gtgaagagcg gcaagctttg gttaatcagt taaaatcagc gaagactaca gtaataagcc 600
 ttggtcagga gctgcaaaaac gaaaaaaaac ttgctgaaga tctcaaatth gagatcaagg 660
 gccttcaaaa tgacctcatg aatacgaagg aggataagaa gaaattgcag gaagagctta 720
 aagagaagct tgatttgatt caagttcttg aagaaaagat tactttgctt actacagaga 780
 tcaaagataa agaggtgagt cttcggagta acacctctaa actagctgaa aaagaatcgg 840
 aggtaaatag tttgagcgat atgtatcaac aatcccagga tcagctgatg aatttgactt 900
 cagagatcaa agaacttaaa gatgaaatcc agaaaagaga gagagaactg gagttgaaat 960
 gtgtatcaga agacaacctg aatgtgcaat taaattcttt gtcctcga 1009

<210> 11
 <211> 1103

<212> DNA

<213> *Nicotiana tabacum*

<400> 11

cttgagaaat	caagaagtga	agcttctgat	atagtagaac	aactacagca	gtcgaaggcat	60
ctttgctcta	agcttgaagc	tgaggtttct	aagcttcaga	tggaattgga	ggaaacaagg	120
acattgttac	agaagaacat	tgatgagaca	aaacgtggtg	cagagttctt	agctgcggag	180
ctgaccacta	ctagggagct	tctaaagaaa	acaaatgaag	aaatgcacac	catatccaat	240
gaactagctg	ctgttactga	aaatcgtgat	aacttacaga	cggagctagt	tgatgtctac	300
aagaaagcag	aacgtgctgt	taatgaactg	aaacaagaaa	agaatattgt	cgtgacattg	360
gagaaagagc	taacattttt	ggaggctcaa	attacaagag	agaaagagtc	acggaagaat	420
ctggaagaag	agttggaaaag	ggctacagaa	tcacttgatg	agatgaacag	aaatgctttt	480
gcacttgcaa	aggagctgga	gctcgctaata	tctcgtatct	ctagcctcaa	agacgagaga	540
gaagtgtccc	aaaagtctgt	ttctgagcag	aagcaaattt	ctcaagaagc	ccgagaaaac	600
cttgaagatg	cccatagcct	ggtgatgaaa	cttggcaagg	aacgcgagag	tctggagaag	660
agagcaaaga	aattggaaga	tgaaatggca	tcagcaaaag	gtgagatttt	gcggttgccg	720
acacaagtaa	attcggtaaa	agctcctggt	aacaaagagg	aaaaagttga	agctggggaa	780
aaggcaacag	taacagtga	gagaacaacc	aggaggagga	agactgctac	tcctgcttct	840
caacaagaag	gctcataatt	tgctgtttct	gaagtgcacat	atatccttcc	ttttttcctt	900
gactcatatt	aattgcaacg	agggtagatt	attggttcat	tatataaaac	cagaatgagg	960
atattgcctt	tgtaagaaac	tttctgcaag	ctgtattctc	agtgaagtaa	tttccaggcg	1020
agaagttgcc	caaataaatg	agatattatt	gttgcaagta	ccaaatttgg	aagggattgt	1080
ttgatatcaa	aaaaaaaaaa	aaa				1103

<210> 12

<211> 912

<212> DNA

<213> *Nicotiana tabacum*

<400> 12

atgagacaaa	acgtggtgca	gagctcttag	ctgcggagct	gaccactact	agggagcttc	60
taaagaaaaac	aaatgaagaa	atgcacacta	tgtctcatga	actagcggct	gttactgaaa	120
attgtgataa	cttacagacg	gagctagtgt	atgtctacaa	gaaagcagaa	cgtgctgctg	180
atgaactgaa	acaagaaaag	aatattgtcg	tgacactgga	gaaagagcta	acatttttgg	240
aggctcaaat	tacaagagag	aaagagtcac	ggaagaatct	ggaagaagag	ctggaaaagg	300
ctacggaatc	acttgatgag	atgaaccgaa	atgcttttgc	acttgcaaag	gagcttgagc	360
ttgctaattc	tcatatttct	agcctcgagg	atgagagaga	agtgtccaa	aagtctgttt	420
ctgagcagaa	acaaatttct	caagaatccc	gagaaaacct	tgaagatgcc	catagcctgg	480
taatgaaact	tggaagga	cgcgagagtc	tggaagaag	agcaaagaaa	ttggaagatg	540
aaatggcatc	agcaaaagg	gagattttgc	ggctgcggac	ccaagtaaat	tcggtaaaag	600
ctcctgttaa	caatgaggaa	aaagttgaag	ctggggaaaa	ggcagctgta	acagtgaaga	660
gaaccaggag	gaggaagact	gctactcagc	ctgcttctca	gcaagaaagc	tcatagtttg	720
ctgttctaaa	gtgacatata	tttccttttt	gtccttgact	caaattgatt	gcgacgagaa	780
tagattaatg	gtgtattata	gagaagccag	aattaggata	ttgcccttgt	aagaaacttc	840
ctgcaagctg	tattctcagt	gagtgatat	ttccagggtga	gaagttgcac	aaacaaaaaa	900
aaaaaaaaaa	aa					912

<210> 13

<211> 905

<212> DNA

<213> *Nicotiana tabacum*

<400> 13

cgagatgtga	atcagaagac	aacctgaatg	tgcaattaaa	ttctttgctc	gttgagagag	60
atgaatctaa	aaaagagctt	gatgctattc	aaaaggaata	cagcgagttc	aagtccattt	120
cgagaagag	agtggcttca	gatgccaaag	tggtggggga	acaagaaaag	agactacacc	180
agctcgagga	acaacttggt	actgccgtaa	gtgaagtgaag	aaaaataaaa	gtgctaattg	240
ctaatttgac	tcaagcaaaa	gaaaacctaa	ggagaatgct	ggacgctgag	ctggaaaatg	300
taagcaagtt	gaagctagag	gtccagggtta	ctcaggaaac	tcttgagaaa	tcaagaagtg	360
aagcttctga	tatagtagaa	caactacagc	agtcgaggca	tctttgctct	aagcttgaag	420
ctgaggtttc	taagcttcag	atggaattgg	aggaaacaag	gacattgtta	cagaagaaca	480
ttgatgagac	aaaacgtgg	gcagagctct	tagctgcgga	gctgaccact	actagggagc	540
ttctaaagaa	aacaaatgaa	gaaatgcaca	ccatatccaa	tgaactagct	gctgttactg	600
aaaatcgtga	taacttacag	acggagctag	ttgatgtcta	caagaaagca	gaacgtgctg	660

ttaatgaact	gaaacaagaa	aagaatattg	tcgtgacatt	ggagaaagag	ctaacatttt	720
tggaggctca	aattacaaga	gagaaagagt	caccgaagaa	tctggaagaa	gagttggaaa	780
gggctagctc	gcttaagtac	aggagatgga	gaatccaccg	aagaatgaag	tagtggcaga	840
tcatctgcgt	ccaagcaagt	tacttcacca	acagaaaact	tggatttgta	cctgcctgct	900
ctccg						905

<210> 14
 <211> 1597
 <212> DNA
 <213> *Nicotiana tabacum*

<400> 14						
cggcctctga	aatcttcttc	tttttatcac	tttcggagtg	gaaatcggga	gaaaccaacc	60
aactttgtaa	tggggaggtt	ttgttttccc	caatctccac	tctctcattc	tctcttttct	120
tcttcatcaa	tatcttcttc	ccaatttaca	cccttgcttt	tttccccaag	aaatgcgcaa	180
aaatgtaaaa	agaaaaatgcc	agctatggca	tgtatacact	cggagaatca	aaaggaaaagc	240
gaattctgca	gcagaagaac	gattcttttc	gtgggtttct	ctgttcttcc	acttctcagc	300
ttgagggcaa	atgcttttga	aggcttgctc	gtagattctc	aagtaaaagc	acagccgcag	360
aaagaggaga	cgagcaacaa	tccaaggaaa	tgcagagaat	cccttctttt	ctctacttaa	420
tggacttgga	gttttttggt	caggcgtgct	tggttctctt	tatgccttgg	ctcgaaacga	480
gaaggccgtt	tctgatgcaa	ccattgaatc	tatgaaaaat	aagctgaagg	agaaagaagc	540
cacattcgtt	tcatggagaa	gaaattccag	tctgagctgc	tgaacgaaag	ggatatacga	600
ataaatcaac	ttaagagggc	aggcgaagaa	cggcaagctc	tggttaacca	attgaattca	660
gcaaagagta	cagtaactaa	ccttggtcag	gagctgcaaa	aagaaaaacg	aattgctgaa	720
gagctcatag	ttcagatcga	gggccttcaa	aataacctca	tgcagatgaa	ggaggataag	780
aaaaaattgc	aggaggagct	taaagagaag	cttgatttga	tacaagttct	gcaagaaaaag	840
ataactttac	ttactacaga	gatcaaagat	aaagaggcat	ctcttcagag	tacaacctct	900
aaactagctg	aaaaagaatc	agaggtagat	aaattgagct	caatgtatca	ggaatcccag	960
gatcagctga	tgaatttgac	ttcagaaatc	aaagaactta	aagtcgaagt	ccagaaaaga	1020
gagagagaac	tagagttgaa	acgtgaatca	gaagacaacc	ttaatgtgcg	attaaattct	1080
ttgctcgttg	agagagatga	atctaaaaaa	gagcttgatg	ctattcaaaa	ggaatacagc	1140
gagttcaagt	ccatttcaga	gaagaaaagt	gcttctgatg	ccaagctgtt	gggggaacaa	1200
gaaaagagac	tacaccagct	cgaggaacaa	cttggcactg	cctcagatga	agtacgcaaa	1260
aataatgtgc	taatcgctga	tctgactcaa	gaaaaagaaa	acttaaggag	aatgctggac	1320
gctgagctgg	aaaacataag	caagttgaag	ctagaggtcc	aggttactca	ggaaactctt	1380
gagaaatcta	gaagtgatgc	ttctgatata	gcacaacaac	tacagcagtc	gaggcatctt	1440
tgctctaagc	ttgaagctga	ggtttctaaa	cttcagatgg	aattggagga	aacaagaaca	1500
tcattacgga	ggaacattga	tgagacaaaa	cgtggtgcag	agctcttagc	tgcggagctg	1560
accactacta	gggagcttct	aaagaaaaaa	aaaaaag			1597

<210> 15
 <211> 564
 <212> DNA
 <213> *Nicotiana tabacum*

<400> 15						
gaggaacaac	ttggcactgc	ctcagatgaa	gtacgcaaaa	ataatgtgct	aatcgctgat	60
ctgactcaag	aaaaagaaaa	cttaaggaga	atgctggacg	ctgagctgga	aaacataagc	120
aagttgaagc	tagaggtcca	ggttactcag	gaaactcttg	agaaatctag	aagtgatgct	180
tctgatatag	cacaacaact	acagcagctg	aggcatcttt	gctctaagct	tgaagctgag	240
gtttctaaac	ttcagatgga	attggaggaa	acaagaacat	cattacggag	gaacattgat	300
gagacaaaac	gtggtgcaga	gctcttagct	gcggagctga	ccactactag	ggagcttcta	360
aagaaaacaa	atgaagaaat	gcacactatg	tctcatgaac	tagcggctgt	tactgaaaat	420
tgtgataact	tacagacgga	gctagttgat	gtctacaaga	aagcagaacg	tgctgctgat	480
gaactgaaac	aagaaaagaa	tattgtcgtg	acactggaga	aagagctaac	atttttggag	540
gctcaaatta	caagagagaa	agag				564

<210> 16
 <211> 2154
 <212> DNA
 <213> *Lycopersicon esculentum*

<400> 16
 atggcaactt cttgttttcc tccattttct gcttcattctt cttcattatg ttctttcccaa 60
 tttaacacct tgctttcttg cccaagaaat acccaaatat gtagaaagaa gagaccggtt 120
 atggcgagta tgcactcgga aaatcaaaaag gaaagtaatg tctgcaacag aagatcgatt 180
 ctatttgtgg gattctcagt tcttccactt ctcaatttga gggcaagagc tctcgaaggc 240
 ttgtcaacag attctcaagc acagccgcag aaagaggaaa ccgagcaaac aatccaagga 300
 agtgcaggga atcccttcgt ttctctactt aatggacttg gtgttgttg ttcaggcgtg 360
 cttggttctc tttatgcctt ggctcgaaat gagaaggcag ttccagatgc aaccattgaa 420
 tctatgaaaa ataagctgaa ggacaaggaa gatgcatttg tttcaatgaa gaagcaattt 480
 gagtccgaat tgctgagcga aagggaagat cgaaataagc taattaggcg agaaggtgaa 540
 gagcggaag ctttggttaa tcagttaaaa tcagcgaaga ctacagtaat aagccttggt 600
 caggagctgc aaaacgaaaa aaaacttgct gaagatctca aatttgagat caagggcctt 660
 caaaatgacc tcatgaatac gaaggaggat aagaagaaat tgcaggaaga gcttaaagag 720
 aagcttgatt tgattcaagt tcttgaagaa aagattactt tgcttactac agagatcaaa 780
 gataaagagg tgagctctcg gagtaacacc tctaaactag ctgaaaaaga atcggaggta 840
 aatagtttga gcgatatgta tcaacaatcc caggatcagc tgatgaattt gacttcagag 900
 atcaaagaac ttaaagatga aatccagaaa agagagagag aactggagtt gaaatgtgta 960
 tcagaagaca acctgaatgt gcaattaaat tctttgctcc tcgagagaga tgaatctaaa 1020
 aaagagcttc atgctattca aaaggaatac agtgagttca agtccaattc tgatgagaag 1080
 gtggcttcag atgcgaagct gttgggggaa caagagaaga gactacacca gcttgaggaa 1140
 caacttgga ctgccttaag tgaagcaagt aaaaatgaag tgctaattgc tgatctgact 1200
 cgagaaaaag aaaaccttag gagaatggtg gatgctgagc tggacaatgt aaacaagtta 1260
 aagcaagaga ttgaagtcac tcaggaaagt cttgagaatt caagaagtga agtttctgat 1320
 ataacagtac aactagagca gttgagggat ctttgctcca aacttgaagc tgaggtttct 1380
 aaacttcaga tgggaattgga ggaaacaagg gcatcattac agaggaacat tgatgaaaca 1440
 aaacacagtt cacagctctt agctgctgag ttgaccacta ctaaggagct tctaaagaaa 1500
 acaaatgaag aaatgcacac tatgtcagat gaactagtag ctgtttctga aaatcgtgat 1560
 agcttacaga cacagctagt tgatgtctat aagaaagcag aacatactgc taatgaactg 1620
 aaacaagaaa agagcattgt tgcaacacta gaagaagagt taaaatttct ggagtctcaa 1680
 attacacgag agaaagagtt acggaagagt ctggaagacg agttagaaaa ggctacagaa 1740
 tctcttgatg agattaaccg aaatgtgttg gcacttgacg aggagctgga gcttgctact 1800
 tctcgtaatt ctagcctcga agacgagaga gaagtgtctc gacagtctgt ttctgagcag 1860
 aagcaaatgt cacaagaagc ccaagaaaat ctggaagacg cccatagcct ggtgatgaaa 1920
 cttggcaagg aacgcgaaag tcttgagaag agagcaaaga aattggaaga tgaaatggca 1980
 gcagcaaaaag gtgagatttt gcggctacgg agccaaataa actcagtaaa agctccagtg 2040
 gaggatgagg aaaaagttgt tgctggggaa aaggaaaagg tgaaggcaac agtaacagca 2100
 aagaaaacta ccaggagaag gaagagtgtc actgttaagc aagaggaacc ctag 2154

<210> 17
 <211> 717
 <212> PRT
 <213> *Lycopersicon esculentum*

<400> 17
 Met Ala Thr Ser Cys Phe Pro Pro Phe Ser Ala Ser Ser Ser Ser Leu
 1 5 10 15
 Cys Ser Ser Gln Phe Thr Pro Leu Leu Ser Cys Pro Arg Asn Thr Gln
 20 25 30
 Ile Cys Arg Lys Lys Arg Pro Val Met Ala Ser Met His Ser Glu Asn
 35 40 45
 Gln Lys Glu Ser Asn Val Cys Asn Arg Arg Ser Ile Leu Phe Val Gly
 50 55 60
 Phe Ser Val Leu Pro Leu Leu Asn Leu Arg Ala Arg Ala Leu Glu Gly
 65 70 75 80
 Leu Ser Thr Asp Ser Gln Ala Gln Pro Gln Lys Glu Glu Thr Glu Gln
 85 90 95
 Thr Ile Gln Gly Ser Ala Gly Asn Pro Phe Val Ser Leu Leu Asn Gly
 100 105 110

Leu Gly Val Val Gly Ser Gly Val Leu Gly Ser Leu Tyr Ala Leu Ala
 115 120 125
 Arg Asn Glu Lys Ala Val Ser Asp Ala Thr Ile Glu Ser Met Lys Asn
 130 135 140
 Lys Leu Lys Asp Lys Glu Asp Ala Phe Val Ser Met Lys Lys Gln Phe
 145 150 155 160
 Glu Ser Glu Leu Leu Ser Glu Arg Glu Asp Arg Asn Lys Leu Ile Arg
 165 170 175
 Arg Glu Gly Glu Glu Arg Gln Ala Leu Val Asn Gln Leu Lys Ser Ala
 180 185 190
 Lys Thr Thr Val Ile Ser Leu Gly Gln Glu Leu Gln Asn Glu Lys Lys
 195 200 205
 Leu Ala Glu Asp Leu Lys Phe Glu Ile Lys Gly Leu Gln Asn Asp Leu
 210 215 220
 Met Asn Thr Lys Glu Asp Lys Lys Lys Leu Gln Glu Glu Leu Lys Glu
 225 230 235 240
 Lys Leu Asp Leu Ile Gln Val Leu Glu Glu Lys Ile Thr Leu Leu Thr
 245 250 255
 Thr Glu Ile Lys Asp Lys Glu Val Ser Leu Arg Ser Asn Thr Ser Lys
 260 265 270
 Leu Ala Glu Lys Glu Ser Glu Val Asn Ser Leu Ser Asp Met Tyr Gln
 275 280 285
 Gln Ser Gln Asp Gln Leu Met Asn Leu Thr Ser Glu Ile Lys Glu Leu
 290 295 300
 Lys Asp Glu Ile Gln Lys Arg Glu Arg Glu Leu Glu Leu Lys Cys Val
 305 310 315 320
 Ser Glu Asp Asn Leu Asn Val Gln Leu Asn Ser Leu Leu Leu Glu Arg
 325 330 335
 Asp Glu Ser Lys Lys Glu Leu His Ala Ile Gln Lys Glu Tyr Ser Glu
 340 345 350
 Phe Lys Ser Asn Ser Asp Glu Lys Val Ala Ser Asp Ala Lys Leu Leu
 355 360 365
 Gly Glu Gln Glu Lys Arg Leu His Gln Leu Glu Glu Gln Leu Gly Thr
 370 375 380
 Ala Leu Ser Glu Ala Ser Lys Asn Glu Val Leu Ile Ala Asp Leu Thr
 385 390 395 400
 Arg Glu Lys Glu Asn Leu Arg Arg Met Val Asp Ala Glu Leu Asp Asn
 405 410 415
 Val Asn Lys Leu Lys Gln Glu Ile Glu Val Thr Gln Glu Ser Leu Glu
 420 425 430
 Asn Ser Arg Ser Glu Val Ser Asp Ile Thr Val Gln Leu Glu Gln Leu
 435 440 445

Arg Asp Leu Cys Ser Lys Leu Glu Ala Glu Val Ser Lys Leu Gln Met
 450 455 460
 Glu Leu Glu Glu Thr Arg Ala Ser Leu Gln Arg Asn Ile Asp Glu Thr
 465 470 475 480
 Lys His Ser Ser Glu Leu Leu Ala Ala Glu Leu Thr Thr Thr Lys Glu
 485 490 495
 Leu Leu Lys Lys Thr Asn Glu Glu Met His Thr Met Ser Asp Glu Leu
 500 505 510
 Val Ala Val Ser Glu Asn Arg Asp Ser Leu Gln Thr Glu Leu Val Asp
 515 520 525
 Val Tyr Lys Lys Ala Glu His Thr Ala Asn Glu Leu Lys Gln Glu Lys
 530 535 540
 Ser Ile Val Ala Thr Leu Glu Glu Glu Leu Lys Phe Leu Glu Ser Gln
 545 550 555 560
 Ile Thr Arg Glu Lys Glu Leu Arg Lys Ser Leu Glu Asp Glu Leu Glu
 565 570 575
 Lys Ala Thr Glu Ser Leu Asp Glu Ile Asn Arg Asn Val Leu Ala Leu
 580 585 590
 Ala Glu Glu Leu Glu Leu Ala Thr Ser Arg Asn Ser Ser Leu Glu Asp
 595 600 605
 Glu Arg Glu Val Leu Arg Gln Ser Val Ser Glu Gln Lys Gln Ile Ser
 610 615 620
 Gln Glu Ala Gln Glu Asn Leu Glu Asp Ala His Ser Leu Val Met Lys
 625 630 635 640
 Leu Gly Lys Glu Arg Glu Ser Leu Glu Lys Arg Ala Lys Lys Leu Glu
 645 650 655
 Asp Glu Met Ala Ala Ala Lys Gly Glu Ile Leu Arg Leu Arg Ser Gln
 660 665 670
 Ile Asn Ser Val Lys Ala Pro Val Glu Asp Glu Glu Lys Val Val Ala
 675 680 685
 Gly Glu Lys Glu Lys Val Lys Ala Thr Val Thr Ala Lys Lys Thr Thr
 690 695 700
 Arg Arg Arg Lys Ser Ala Thr Val Lys Gln Glu Glu Pro
 705 710 715

<210> 18
 <211> 407
 <212> DNA
 <213> Nicotiana tabacum

<400> 18
 tcgaggaaca acttggcact gcctcagatg aagtacgcaa aaataatgtg ctaatcgctg 60
 atctgactca agaaaaagaa aacttaagga gaatgctgga cgctgagctg gaaaacataa 120
 gcaagttgaa gctagaggtc caggttactc aggaaactct tgagaaatct agaagtgatg 180
 cttctgatat agcacaacaa ctacagcagt cgaggcatct ttgctctaag cttgaagctg 240
 aggtttctaa acttcagatg gaattggagg aaacaagaac atcattacgg aggaacattg 300
 atgagacaaa acgtggtgca gagctcttag ctgcggagct gaccactact agggagcttc 360
 taaagaaaaa aaaaaaagga attcctgcag cccgggggat ccactag 407

<210> 19
 <211> 1491
 <212> DNA
 <213> Glycine max

<400> 19
 gtgatgtcat ggagaaggaa tacaatgata taaagttagc tgctgtaaag aaggctgctt 60
 tggactctaa gggttttaaga gaaaaagaag aggagcttca tcagctaaag gatcagtttg 120
 aacttgccct aggtgaagca agcaaaagcc agatcgatcat tgctgattta tcccaacaaa 180
 gagatgattt gaaggaggct ctagataatg aatctagcaa ggtgaatcat ttgaagcaag 240
 aactccaagt taccctggag aatcttgcaa aatcaagaaa tgagtctgct gaattggaaa 300
 accttctaac tttgtcaaac aaactgtgca aagagctcga gctcgaggct tctaagctct 360
 catctgagct cactgaggtt aatgaatcgc tacagagaaa ccttgatgat gcgaaacatg 420
 aggcagaaat gctagcaagt gagcttataa ctgccaagga acacttgaag gaagcacaag 480
 cagagctgca aggttgtcaa aagaatctga cagctgctct tgaaaagaat gatagcctac 540
 aaaaagaatt agttgaagtc taaaaaagg ctgaaagcac agcagaggat ttgaaggaac 600
 aaaaacagtt agttgcttct ctgaacaaag atttacaagc attagagcag caagtctcaa 660
 aagacaagga gtcccgaaaa tctcttgaga gggacctgga ggaggcgacc atatcactag 720
 atgaaatgaa ccgaaatgag gtgatccttt ctggggaaact acagagagct aattctcttg 780
 tttctagcct tgaaaaagag aaagatgtgc ttattaagtc cctaaccaac caaagaaatg 840
 catgcaaaga ggccaagac aacattgaag atgctcataa ccttatcatg aaacttggca 900
 aagaaagaga gaatttagag aaaaaaggta agaaatttga agaggaattg gcttctgcca 960
 aggggtgagat attgcgcttg aagagtcgaa tcaattcttc aaaagttgct gttaacaatg 1020
 gccagtgca gaaagatgga ggtgaaaaaa aggtcaaccc ttcaaaagtt gcggtaaaca 1080
 atgagcaagc acagaaagat gaagggtgaaa acaaggttac tgtaagtgca cggaagactg 1140
 tcagaagaag aaaggctaatt ccacaataac agagaaatta gagagttttc tattaaaaat 1200
 attccgatta ggatcatgat attctgtaatt aaactatttg gaagccagtt gattctattc 1260
 acttttggca tgcaaatatt ttcattgtttt gcaatagtat tgacaaatta aatgacactg 1320
 taggaattgt taagctaagc tttttggaga gttgatttct gatagtaaac ctaaaaaaa 1380
 aggttaagaat actattacca accttagtct gcaacattat acattagtgt atatacagct 1440
 acttttccat gtctatgaag caaatcgaca agcttggtgc caaaaaaaa a 1491

<210> 20
 <211> 388
 <212> PRT
 <213> Glycine max

<400> 20
 Asp Val Met Glu Lys Glu Tyr Asn Asp Leu Lys Phe Ser Ala Val Lys
 1 5 10 15
 Lys Ala Ala Leu Asp Ser Lys Val Leu Arg Glu Lys Glu Glu Leu
 20 25 30
 His Gln Leu Lys Asp Gln Phe Glu Leu Ala Leu Gly Glu Ala Ser Lys
 35 40 45
 Ser Gln Ile Val Ile Ala Asp Leu Ser Gln Gln Arg Asp Asp Leu Lys
 50 55 60
 Glu Ala Leu Asp Asn Glu Ser Ser Lys Val Asn His Leu Lys Gln Glu
 65 70 75 80
 Leu Gln Val Thr Leu Glu Asn Leu Ala Lys Ser Arg Asn Glu Ser Ala
 85 90 95
 Glu Leu Glu Asn Leu Leu Thr Leu Ser Asn Lys Leu Cys Lys Glu Leu
 100 105 110
 Glu Leu Glu Val Ser Lys Leu Ser Ser Glu Leu Thr Glu Val Asn Glu
 115 120 125

Ser Leu Gln Arg Asn Leu Asp Asp Ala Lys His Glu Ala Glu Met Leu
 130 135 140
 Ala Ser Glu Leu Thr Thr Ala Lys Glu His Leu Lys Glu Ala Gln Ala
 145 150 155 160
 Glu Leu Gln Gly Cys Gln Lys Asn Leu Thr Ala Ala Leu Glu Lys Asn
 165 170 175
 Asp Ser Leu Gln Lys Glu Leu Val Glu Val Tyr Lys Lys Ala Glu Ser
 180 185 190
 Thr Ala Glu Asp Leu Lys Glu Gln Lys Gln Leu Val Ala Ser Leu Asn
 195 200 205
 Lys Asp Leu Gln Ala Leu Glu Gln Gln Val Ser Lys Asp Lys Glu Ser
 210 215 220
 Arg Lys Ser Leu Glu Arg Asp Leu Glu Glu Ala Thr Ile Ser Leu Asp
 225 230 235 240
 Glu Met Asn Arg Asn Ala Val Ile Leu Ser Gly Glu Leu Gln Arg Ala
 245 250 255
 Asn Ser Leu Val Ser Ser Leu Glu Lys Glu Lys Asp Val Leu Ile Lys
 260 265 270
 Ser Leu Thr Asn Gln Arg Asn Ala Cys Lys Glu Ala Gln Asp Asn Ile
 275 280 285
 Glu Asp Ala His Asn Leu Ile Met Lys Leu Gly Lys Glu Arg Glu Asn
 290 295 300
 Leu Glu Lys Lys Gly Lys Lys Phe Glu Glu Glu Leu Ala Ser Ala Lys
 305 310 315 320
 Gly Glu Ile Leu Arg Leu Lys Ser Arg Ile Asn Ser Ser Lys Val Ala
 325 330 335
 Val Asn Asn Gly Pro Val Gln Lys Asp Gly Gly Glu Lys Lys Val Asn
 340 345 350
 Pro Ser Lys Val Ala Val Asn Asn Glu Gln Ala Gln Lys Asp Glu Gly
 355 360 365
 Glu Asn Lys Val Thr Val Ser Ala Arg Lys Thr Val Arg Arg Arg Lys
 370 375 380

Ala Asn Pro Gln
385

<210> 21
 <211> 2019
 <212> DNA
 <213> Zea mays

<400> 21
 cggacgcgtg ggcctaaatt tgaagggaca aagggtattg caaaacctga caacactcaa 60
 cctgaaggaa ctcaggctga aactatacct gaagctcgct agcgtgaatc atccttacag 120
 ttggtgcaag aacaacctcc agagaatcca ctgcttggtt ttcttggtat agttggagtt 180
 gctgcctctg gtgttcttgg tgggctgtac ggcacttctc tacaagaaga aaaggccctg 240
 caatcaattg tctcctcaat ggagagcaaa ttggctgaaa atgaggcagc actttcattg 300
 atgagggata attatgagaa acggttactg gaggcagcaag cagcacaaaa gaagcaatct 360
 atgaagttcc aggagcagga agtttctctt tcagggtcagt tggcttcagc aacaaagact 420

```

ttgacatcac  tgagtgaaga  attcagaaag  gagaagaaat  tagctgagga  acttagggat  480
gaaatacaga  gattagagag  tagtatcaca  caagctggca  ttgataatga  tgtgcttgaa  540
actaaattgg  aagaaaagct  tggtagagatt  aattttttgc  aggaaaaggt  aagtttactc  600
aaccaagaaa  ttgatgataa  ggagaagcac  atcaggggaa  tcagtgcac  actttcctcg  660
aaggaagtag  actacaaaaa  gctgaccgct  ttcacaaatc  aaactaaaaa  gagccttgag  720
cttgcaaatt  ctagagtaca  acaactcgag  gaagaactaa  gtacaactaa  gaacgctctc  780
gttttctaaga  tatcttctat  tgattcactc  aatgctaaac  ttgaaacctt  gaactctgaa  840
aagaagaagc  tgacaaaaaa  aataaatgag  ttaatacaag  agtatacaga  cctgaagggt  900
gcttcagaga  caagggcaag  ccatgattcc  aaactactgt  cagaaagaga  tgatctgata  960
aaacagcttg  aggaaaaaact  gtctgttgca  ttaactgatt  ctagcaaaga  tcaagaaaca  1020
attgttgagt  tgaacaagga  gttggatgct  accaaaatga  tgctaaagaa  tgaacttaag  1080
tccatggaag  ctttaaaaga  ttcaattcga  tcatctgaag  aggctctaaa  gacttcaaga  1140
agtgaggttt  ccaaaactttc  caaggagctt  gaggaggcaa  atgaattgaa  tgaggacctg  1200
gtatcacaaa  tttctaaact  ccgagaggaa  tccaatgaaa  tgcaagtaga  tctcactaat  1260
aaactaggag  aggcagaatc  actatctaaa  gctctgtcag  aagatttggc  ttcagtaaat  1320
gaaatgggtc  agaagggaca  agaagaactc  gaagccacct  ctattgagct  ggcactctatt  1380
gctgaagctc  gtgacaactt  gaagaaagaa  ttgctggatg  cgtacaagaa  tttggagtca  1440
accacacatg  agcttgtcga  ggaaagaaaa  attgtgacag  ccttaaacia  ggaacttgaa  1500
gcgttagcga  aacagttgca  gggttgattct  gaagcaagaa  aaagtctcga  atcagacctg  1560
gaggaggcaa  caaagtcact  agatgaaatg  aacaatagcg  cgctgttact  gtctaagaa  1620
cttgagagca  ctcaattctag  gagtgccact  cttgaatctg  agaaggaaat  gctacgcaag  1680
gctctagctg  aacaaacgaa  aatcacaaacc  gaagctaagg  aaaacacaga  ggatgctcag  1740
aaccttatca  caaggcttga  gacagagaag  gagagctttg  aattgaggtg  tagacatctt  1800
gaagaggaat  tggcgttagc  aaaaggtgag  atactgcgcc  taaggaggca  gattagcaca  1860
aacagttctc  agaaaccaag  agcaagagga  ccaccagagg  ccagtgaac  tctgaaggag  1920
caacctgtga  atgattataa  tcagaagacc  agtggagttg  ttgctggaac  tccacagcct  1980
gtgaaaagga  ctgtaaggag  aagaaaaggt  ggcgcataa  2019

```

<210> 22
<211> 672
<212> PRT
<213> Zea mays

<400> 22
Arg Thr Arg Gly Pro Lys Phe Glu Gly Thr Lys Gly Ile Ala Lys Pro
1 5 10 15
Asp Asn Thr Gln Pro Glu Gly Thr Gln Ala Glu Thr Ile Pro Glu Ala
20 25 30
Arg Gln Arg Glu Ser Ser Leu Gln Leu Val Gln Glu Gln Pro Pro Glu
35 40 45
Asn Pro Leu Leu Gly Phe Leu Gly Ile Val Gly Val Ala Ala Ser Gly
50 55 60
Val Leu Gly Gly Leu Tyr Gly Thr Ser Leu Gln Glu Glu Lys Ala Leu
65 70 75 80
Gln Ser Ile Val Ser Ser Met Glu Ser Lys Leu Ala Glu Asn Glu Ala
85 90 95
Ala Leu Ser Leu Met Arg Asp Asn Tyr Glu Lys Arg Leu Leu Glu Gln
100 105 110
Gln Ala Ala Gln Lys Lys Gln Ser Met Lys Phe Gln Glu Gln Glu Val
115 120 125
Ser Leu Ser Gly Gln Leu Ala Ser Ala Thr Lys Thr Leu Thr Ser Leu
130 135 140
Ser Glu Glu Phe Arg Lys Glu Lys Lys Leu Ala Glu Glu Leu Arg Asp
145 150 155 160

Glu Ile Gln Arg Leu Glu Ser Ser Ile Thr Gln Ala Gly Ile Asp Asn
 165 170 175
 Asp Val Leu Glu Thr Lys Leu Glu Glu Lys Leu Gly Glu Ile Asn Phe
 180 185 190
 Leu Gln Glu Lys Val Ser Leu Leu Asn Gln Glu Ile Asp Asp Lys Glu
 195 200 205
 Lys His Ile Arg Glu Leu Ser Ala Ser Leu Ser Ser Lys Glu Val Asp
 210 215 220
 Tyr Gln Lys Leu Thr Ala Phe Thr Asn Gln Thr Lys Lys Ser Leu Glu
 225 230 235 240
 Leu Ala Asn Ser Arg Val Gln Gln Leu Glu Glu Glu Leu Ser Thr Thr
 245 250 255
 Lys Asn Ala Leu Val Ser Lys Ile Ser Ser Ile Asp Ser Leu Asn Ala
 260 265 270
 Lys Leu Glu Thr Leu Asn Ser Glu Lys Lys Lys Leu Thr Lys Lys Ile
 275 280 285
 Asn Glu Leu Ile Gln Glu Tyr Thr Asp Leu Lys Val Ala Ser Glu Thr
 290 295 300
 Arg Ala Ser His Asp Ser Lys Leu Leu Ser Glu Arg Asp Asp Leu Ile
 305 310 315 320
 Lys Gln Leu Glu Glu Lys Leu Ser Val Ala Leu Thr Asp Ser Ser Lys
 325 330 335
 Asp Gln Glu Thr Ile Val Glu Leu Asn Lys Glu Leu Asp Ala Thr Lys
 340 345 350
 Met Met Leu Lys Asn Glu Leu Lys Ser Met Glu Ala Leu Lys Asp Ser
 355 360 365
 Ile Arg Ser Ser Glu Glu Ala Leu Lys Thr Ser Arg Ser Glu Val Ser
 370 375 380
 Lys Leu Ser Lys Glu Leu Glu Glu Ala Asn Glu Leu Asn Glu Asp Leu
 385 390 395 400
 Val Ser Gln Ile Ser Lys Leu Arg Glu Glu Ser Asn Glu Met Gln Val
 405 410 415
 Asp Leu Thr Asn Lys Leu Gly Glu Ala Glu Ser Leu Ser Lys Ala Leu
 420 425 430
 Ser Glu Asp Leu Ala Ser Val Asn Glu Met Val Gln Lys Gly Gln Glu
 435 440 445
 Glu Leu Glu Ala Thr Ser Ile Glu Leu Ala Ser Ile Ala Glu Ala Arg
 450 455 460
 Asp Asn Leu Lys Lys Glu Leu Leu Asp Ala Tyr Lys Asn Leu Glu Ser
 465 470 475 480
 Thr Thr His Glu Leu Val Glu Glu Arg Lys Ile Val Thr Ala Leu Asn
 485 490 495

Lys Glu Leu Glu Ala Leu Ala Lys Gln Leu Gln Val Asp Ser Glu Ala
 500 505 510
 Arg Lys Ser Leu Glu Ser Asp Leu Glu Glu Ala Thr Lys Ser Leu Asp
 515 520 525
 Glu Met Asn Asn Ser Ala Leu Leu Leu Ser Lys Glu Leu Glu Ser Thr
 530 535 540
 His Ser Arg Ser Ala Thr Leu Glu Ser Glu Lys Glu Met Leu Arg Lys
 545 550 555 560
 Ala Leu Ala Glu Gln Thr Lys Ile Thr Thr Glu Ala Lys Glu Asn Thr
 565 570 575
 Glu Asp Ala Gln Asn Leu Ile Thr Arg Leu Glu Thr Glu Lys Glu Ser
 580 585 590
 Phe Glu Leu Arg Cys Arg His Leu Glu Glu Glu Leu Ala Leu Ala Lys
 595 600 605
 Gly Glu Ile Leu Arg Leu Arg Arg Gln Ile Ser Thr Asn Ser Ser Gln
 610 615 620
 Lys Pro Arg Ala Arg Gly Pro Pro Glu Ala Ser Glu Thr Leu Lys Glu
 625 630 635 640
 Gln Pro Val Asn Asp Tyr Asn Gln Lys Thr Ser Gly Val Val Ala Gly
 645 650 655
 Thr Pro Gln Pro Val Lys Arg Thr Val Arg Arg Arg Lys Gly Gly Ala
 660 665 670

<210> 23
 <211> 322
 <212> DNA
 <213> Oryza sativa

<220>
 <223> n= g, a, c or t

<400> 23
 gagagaaaact agttctagga aggacactct tgaagcagag aaaaaaatgt tatcaaaggc 60
 tcttgctgag caacagaaga tcacaactga agctcatgaa aacactgagg atgctcagaa 120
 tcttatctct aggcttcaga ctgagaagga gagttttgaa atgagggcta gacatcttga 180
 agaggagttg gcgtagcaaa aggggtgagat attgcgccta agaaggcaga ttagtacaag 240
 cagatcacag aaagcaaaaa ctcttccaaa cacaaatgca tctccagagg tcagtcaggc 300
 tccangacga gcaggctgtg aa 322

<210> 24
 <211> 107
 <212> PRT
 <213> Oryza alta

<220>
 <223> X= G or R

<400> 24
 Arg Glu Thr Ser Ser Arg Lys Asp Thr Leu Glu Ala Glu Lys Lys Met
 1 5 10 15
 Leu Ser Lys Ala Leu Ala Glu Gln Gln Lys Ile Thr Thr Glu Ala His
 20 25 30

Glu Asn Thr Glu Asp Ala Gln Asn Leu Ile Ser Arg Leu Gln Thr Glu
 35 40 45
 Lys Glu Ser Phe Glu Met Arg Ala Arg His Leu Glu Glu Glu Leu Ala
 50 55 60
 Leu Ala Lys Gly Glu Ile Leu Arg Leu Arg Arg Gln Ile Ser Thr Ser
 65 70 75 80
 Arg Ser Gln Lys Ala Lys Thr Leu Pro Asn Thr Asn Ala Ser Pro Glu
 85 90 95
 Val Ser Gln Ala Pro Xaa Arg Ala Gly Cys Glu
 100 105

<210> 25
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:Primer

<400> 25
 gctctagagg aacaacttgg cactgcc

27

<210> 26
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:Primer

<400> 26
 cgggatcctc ttgtaatttg agcctcc

27